November 2015

Probabilistic Reasoning in Cosmology

Yann Benétreau-Dupin
The University of Western Ontario

Supervisor
Professor Chris Smeenk
The University of Western Ontario

Graduate Program in Philosophy

A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

PROBABILISTIC REASONING IN COSMOLOGY
(Thesis format: Integrated Article)

by

Yann Benétreau-Dupin

Graduate Program in Philosophy

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

Abstract

Cosmology raises novel philosophical questions regarding the use of probabilities in inference. This work aims at identifying and assessing lines of arguments and problematic principles in probabilistic reasoning in cosmology.

The first, second, and third papers deal with the intersection of two distinct problems: accounting for selection effects, and representing ignorance or indifference in probabilistic inferences. These two problems meet in the cosmology literature when anthropic considerations are used to predict cosmological parameters by conditionalizing the distribution of, e.g., the cosmological constant on the number of observers it allows for. However, uniform probability distributions usually appealed to in such arguments are an inadequate representation of indifference, and lead to unfounded predictions. It has been argued that this inability to represent ignorance is a fundamental flaw of any inductive framework using additive measures. In the first paper, I examine how imprecise probabilities fare as an inductive framework and avoid such unwarranted inferences. In the second paper, I detail how this framework allows us to successfully avoid the conclusions of Doomsday arguments in a way no Bayesian approach that represents credal states by single credence functions could.

There are in the cosmology literature several kinds of arguments referring to self-locating uncertainty. In the multiverse framework, different "pocket-universes" may have different fundamental physical parameters. We don't know if we are typical observers and if we can safely assume that the physical laws we draw from our observations hold elsewhere. The third paper examines the validity of the appeal to the "Sleeping Beauty problem" and assesses the nature and role of typicality assumptions often endorsed to handle such questions.

A more general issue for the use of probabilities in cosmology concerns the inadequacy of Bayesian and statistical model selection criteria in the absence of well-motivated measures for different cosmological models. The criteria for model selection commonly used tend to focus on optimizing the number of free parameters, but they can select physically implausible models. The fourth paper examines the possibility for Bayesian model selection to circumvent the lack of well-motivated priors.

Keywords: Cosmology, Cosmological Constant Problem, Measure Problem in Cosmology, Probability, Bayesian Confirmation Theory, Induction, Imprecise Probablities, Indifference, Ignorance, Doomsday Argument, Anthropic Reasoning, Self-Locating Beliefs, Sleeping Beauty Problem, Copernican Principle, Typicality, Bayesian Model Selection, Bayesian Information Criterion, Akaike Information Criterion, Ockham's Razor, Simplicity, Parsimony, Unification

Acknowledgements

A project like this one couldn't have been carried out without the help of many. My research was supported through the Ontario Trillium Scholarship.
I first want to express my sincere gratitude to my supervisor, Chris Smeenk, for his encouragement and intellectual guidance. His mentorship was essential throughout the lengthy process of defining a research project and carrying it to completion. I would like to thank my advisory committee, Wayne Myrvold and Carl Hoefer, for their many enlightening and challenging comments. I owe an immense debt of gratitude to

John Norton for his scholarship and for allowing me to come to Pittsburgh. The expert advice of Jim Joyce, Bill Harper, and Rob Corless, all members of my

thesis committee, will lead me to pursue and widen my research in the future.

A great many other people helped me at various stages of my writing, as teachers, interlocutors, or friends: Melissa, Craig, Molly, Mike, Andy, Josh, Fred, Guillaume, and many others at Western; Nora and Mike at Pitt; Mike Hicks, Casey, Chip, Al Wilson, Erik Curiel, Teddy Seidenfeld, Seamus Bradley, and many conference audience members.

Contents

Abstract **ii**

Acknowledgements **iii**

List of Figures **vi**

List of Appendices **vii**

1 Introduction **1**

1.1 Preliminary Remarks 1

1.1.1 Cosmology as a Science 1

1.1.2 Probabilistic Arguments in Cosmology 5

1.1.3 A Caveat about Cosmological Measures 7

1.1.4 Broader Epistemological Issues 8

1.2 Outline of the Thesis 15

1.2.1 The Bayesian Who Knew Too Much 17

1.2.2 Blurring Out Cosmic Puzzles 19

1.2.3 Lost in the Multiverse 22

1.2.4 Simplicity and Unification in Cosmological Model Selection 25

Bibliography 27

2 The Bayesian Who Knew Too Much **31**

2.1 Introduction: Bayesian Reasoning and Ignorance 31

2.2 A Bayesian Failure? 34

2.2.1 Neutral vs. Disconfirming 34

2.2.2 A Non-Bayesian Notion of Neutral Support? 35

2.3 A Bayesian Notion of Neutral Support 37

2.3.1 Bayesian Credences Need Not Have Sharp Values 37

2.3.2 Neutral Support with Imprecise Credences 39

2.3.3 Norton's Objections 44

2.4 Dissolving Cosmic Confusions 49

2.5 Conclusion 52

Bibliography 53

3 Blurring Out Cosmic Puzzles **55**

3.1 Introduction 55

3.1.1 The Doomsday Argument . . . 55
3.1.2 Anthropic Reasoning in Cosmology . . . 57
3.1.3 The Problem: *Ex Nihilo Nihil Fit* . . . 59
3.2 Typicality, Indifference, Neutrality . . . 60
3.2.1 How Crucial to Those Arguments Is the Assumption of Typicality? 60
3.2.2 A Neutral Principle of Indifference? . . . 62
3.3 Dissolving the Puzzles with Imprecise Credence . . . 63
3.3.1 Imprecise Credence . . . 63
3.3.2 Blurring Out Gott's Doomsday Argument: Apocalypse Not Now . 65
3.3.3 Blurring Out Anthropic Predictions . . . 66
3.4 Conclusion . . . 68
Bibliography . . . 69

4 Lost in the Multiverse 71
4.1 Introduction . . . 71
4.2 Self-Locating Beliefs and Observation Bias . . . 74
4.2.1 The Sleeping Beauty Problem as a Problem about Observation Bias 74
4.2.2 Beauty's Bets in a Rigged Game . . . 77
4.3 Self-Locating Uncertainty and Typicality in Cosmology . . . 78
4.3.1 Typicality Assumptions . . . 78
4.3.2 Confusions about the Sleeping Beauty Problem . . . 85
4.3.3 Anthropic Reasoning . . . 88
4.3.4 Testing the Copernican Principle . . . 96
4.4 Conclusion . . . 101
4.A Appendix A: Beauty's Credence and Betting Behavior . . . 103
4.B Appendix B: Bets and Credence with Many Possible Self-Locations . . . 106
Bibliography . . . 107

5 Simplicity and Unification in Cosmological Model Selection 110
5.1 Introduction . . . 110
5.2 Simplicity and Model Selection . . . 114
5.2.1 Why Simplicity Might Matter . . . 114
5.2.2 Why It Doesn't, and What Model Selection Criteria Require . . . 119
5.3 Limitations of Model Selection Criteria in Cosmology . . . 128
5.4 Measuring Consilience Rather Than Predictive Accuracy or Fit . . . 135
5.5 Conclusions . . . 141
Bibliography . . . 144

6 Concluding Remarks 147
Bibliography . . . 152

A Copyright Permissions 153

B Curriculum Vitæ 158

List of Figures

3.1 Posterior probability distributions for N conditional on r, obtained for $r = 100$ and assuming different likelihood distributions for r conditional on N (i.e., with different assumptions as to our relative place in humanity's timeline), which each peaks at different values $\tau = \frac{r}{N}$. The lowermost curve corresponds to a likelihood distribution that peaks at $\tau \to 0$, i.e., if we assume $N \to \infty$. 62

4.1 Illustration of the drift δz over δt, from (Uzan, 2010, § 1.2.4) 100

5.1 Different polynomial curves of different orders (from 1, linear, to 6). The higher the order, the better the curve fits the data; only the 6th-order curve fits all the data points. 116

5.2 Two estimates (in red and in black) of the WMAP angular power spectrum of the Cosmic Microwave Background radiation. The continuous lines are the best-fitting models from inflationary cosmology for each estimate. Source: NASA/WMAP. . . . 135

List of Appendices

Copyright Permissions 153
Curriculum Vitæ 158

Chapter 1

Introduction

1.1 Preliminary Remarks

Probabilistic arguments are a crucial but problematic ingredient of contemporary cosmology. They have been used, e.g., to make predictions regarding the values of fundamental constants by the theory of eternal inflation. Cosmology raises novel philosophical questions regarding, for instance, the status of arguments from fine-tuning, the explanatory strength of appeal to typicality for a unique system, the role of self-locating uncertainty, and the probative value of anthropic considerations. These problems are receiving increasing attention among cosmologists, and only recently have philosophers turned to these topics. All of these problems are connected in the sense that they involve assignments of probability, and as a consequence they need to be addressed together as problems regarding the role and nature of probabilities in cosmology. My research in this regard aims at identifying and assessing lines of arguments and problematic principles in probabilistic reasoning in cosmology.

1.1.1 Cosmology as a Science

Speaking of probabilities in cosmology may seem to be problematic in the same way that cosmology *as a science* seems problematic: there is only *one* universe, and as a consequence one can wonder whether it makes sense to speak of laws of the universe, and how to justify probability values when we cannot measure frequencies. The scientific character of cosmology is not a topic I address in any of the chapters in this work. However, it bears saying a few preliminary remarks about this issue, so as to explain in what way I see cosmology as an interesting context for the kind of broader questions I touch on.

The universe as a whole is a peculiar object, comprising all the physical world, self-contained, and unique. Skepticism as to the possibility to establish a science of the universe considered as a whole usually rests on several lines of argument, including:

- the difficulty to define its object; see, e.g., Roberto Torretti's argument in (Torretti, 2000) that cosmology lacks a satisfying theoretical framework because of the incompatibility between general relativity and quantum mechanics (even the account of the Cosmological Microwave Background radiation requires both theories),

- the difficulty to test our models (particularly because we only have access to a very small portion of the universe) and the role of fundamental and apparently a priori assumptions (such as the Copernican principle),

- the legitimacy to extrapolate local laws to the totality of the universe,

- the possibility that there be a physical science—i.e., a science aimed at discovering physical laws—of a unique object.

All this leads many to question the scientific character of cosmology, or even the mere possibility that there can be a science of the universe (at least a nomothetic science of the universe). The following concern, indeed, is almost commonplace:

> It is very questionable whether the study of any phenomenon that is not repeatable can call itself a science at all. It would be sad however to abandon the whole fascinating area to the priesthood. But if we are going to lend this unique subject any kind of scientific respectability we have to look at all

> its claims with a great circumspection and listen to its proponents with even greater scepticism than is usually necessary. (Disney, 2000, 1126)

This criticism echoes claims made by, e.g., Munitz (1952, 1962, 36) or Ellis (2007), according to which the uniqueness of its object sets cosmology apart from all other physical disciplines, and prevents it from seeking physical laws. Ellis formulated the argument behind such a claim as follows:

Thesis 1 The universe itself cannot be subjected to physical experimentation. (. . .)

Thesis 2 The universe cannot be observationally compared with other universes. (. . .)

Thesis 3 (consequence of 1 & 2) The concept of 'Laws of Physics' that apply to one object only is questionable. *We cannot scientifically establish 'laws of the universe' that might apply to the class of all such objects, for we cannot test any such proposed law except in terms of being consistent with one object (the observed universe).* (Ellis, 2007, 1216–1217, emphasis mine)

On such an account, if the uniqueness of the universe may preclude the scientific character of cosmology, it is because our usual distinctions between laws (necessary) and initial conditions (contingent) don't apply in cosmology, since the initial conditions of the universe are unchangeable, given only once and for all. This problem, Ellis acknowledged, is not specific to cosmology, which he likens to historical or geographical sciences whose aim is to describe an object and its evolution, rather than to find out its laws.

The issue of the uniqueness of the universe as a primary challenge to the scientific character of cosmology can illustrate how relevant cosmology can be to philosophy, and philosophy to cosmology. Indeed, if what is at stake is the possibility of cosmology as a branch of physics, then addressing this philosophical issue matters to cosmology. On the other hand, if what is at stake is the nature of a physical law and whether it presumes a distinction between the contingent and the necessary in the physical facts, then cosmology will provide us with an ideal case study.

Yet it has been argued that asking whether the uniqueness of the universe poses a challenge to the scientific character of cosmology rests on a flawed conception of what it means to be a physical law. Ellis's or Munitz's conception of physical laws usually implies that laws need to apply to multiple instances of a phenomenon given to us completely (and not in part) and to a set of objects that have the same underlying behavior. Cosmological models are usually based on Einstein field equations (EFE), which express, at the largest scale, a relationship between the universe's spatial curvature, its density, and its distribution of matter. EFE are of the following form:

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu} = -\frac{8\pi G}{c^4}T_{\mu\nu} + \Lambda g_{\mu\nu}$$

where $R_{\mu\nu}$ and R are, respectively, the Ricci tensor and Ricci scalar, $g_{\mu\nu}$ is the metric tensor, G is Newton's constant, $T_{\mu\nu}$ is the stress-energy tensor, and Λ is the cosmological constant. Now, as argued in (Smeenk, 2008, 2013, 626), if EFE are taken to be cosmological laws, then every sub-region of a solution of an EFE is also a solution to the same equation. Thus, we could identify a multiplicity of instances to which the global law applies. Therefore, Munitz's and Ellis's claim that cosmology cannot aim at finding laws of the universe because such laws cannot apply to a multiplicity of instances fails for the EFE.

Moreover, the requirement of a distinction between law and instance is problematic insofar as it considers that events or phenomena are instances of a law. However, "[t]he motion of Mars is not an "instance" of Newton's laws; rather, the motion of Mars is well approximated by an equation derived from Newton's laws along with a number of other assumptions." (Smeenk, 2013, 626) The idea that the phenomena can constitute "instances" of laws rests on the idea that physical laws can completely capture the phenomena, which comes to seeing laws and phenomena as both having the same empirical content. But physical laws do not completely describe phenomena; rather, when initial conditions are given, they enable us to give a description that accurately represents the phenomena, and allow further refinements.

The uniqueness of the universe does represent an empirical and conceptual challenge to its scientific study, but not an impossibility in principle. More generally, because of the nature of its object of study, cosmology *exacerbates* difficulties that are common to many physical domains. Likewise, many of the problems and conclusions of the four papers in this dissertation are not specific to cosmology, but they are particularly relevant to cosmology.

1.1.2 Probabilistic Arguments in Cosmology

Probabilistic arguments in cosmology are used for confirming cosmological models, for making predictions about the value of physical parameters, and even for motivating research programs. For instance, three chapters in this dissertation especially deal with probabilistic arguments that appeal to anthropic considerations in order to solve the cosmological constant problem (see below §§ 1.2.1, 2). Likewise, probabilistic arguments are at the basis—explicitly or not—of fine-tuning claims that motivate theories of the initial state of the universe. The cosmological standard model, which gives us a description of global properties of the universe, leads to singularities when extrapolated backward in time. Moreover, it leads to an extremely fine-tuned "initial state" of the universe. Hot big bang cosmology in effect requires that the early universe be highly homogeneous in spite of the fact that separated regions were causally disconnected, and it requires that the global topology of the early universe be extremely close to flatness. Some see this fine-tuning as a problem, and refer to these two requirements as the "horizon problem" and the "flatness problem", respectively. Alan Guth, most notably, argued that such fine-tuning cries out for an explanation, and he proposed in (Guth, 1981) that a stage of exponential expansion in the early universe (right after recombination, around 10^{-35}s after the initial singularity) could solve both problems.

Whether or not inflation does solve the problems it claims to solve,[1] the view that there are problems to be solved rests on an argument from fine-tuning, which implies a probabilistic judgment. In other words, appealing to the initial flatness and homogeneity of the early universe to motivate inflation rests on a claim—often made very explicitly in the literature—that such initial conditions are extremely improbable or extremely unlikely, and therefore implausible and unphysical. One might wonder *on what measure* such a state is improbable (see below § 1.1.3), but it is usually understood that however we choose that measure, the "initial states" selected by compatibility with observations must be extremely close to the maximally symmetric Friedman-Lemaître-Robertson-Walker (FLRW) models,[2] and therefore, presumably, an extremely small subset of the space of solutions of EFE.

On the other hand, one can wonder why we should expect the initial state not to be special. Furthermore, as argued in, e.g., (Feynman, 1967; Penrose, 1979), considerations from statistical mechanics would lead us to expect that the initial state *should be extremely improbable.* This line of reasoning dates back from the work of Ludwig Boltzmann and the search for a justification for time's arrow. On a Boltzmannian account of statistical mechanics and the approach to equilibrium, a system's entropy is as likely to increase toward the future (and time's arrow to flow in the right direction) as it is likely to increase *toward the past*, unless we can show or posit that it was lower in the past. According to David Albert, all posits of a "uniform-over-the-present-macrocondition distribution" used in some Boltzmannian accounts are thus "bound to fail—unless they concern nothing less than the *entirety* of the universe at nothing later than its *beginning.*" (Albert, 2000, 82,85)

Eventually, the status of this account of time's arrow may decide whether or not a rationale for inflationary cosmology—but also, eternal inflation and predictions in the

[1] See (Smeenk, 2014; McCoy, 2015) for discussions.

[2] Friedman-Lemaître-Robertson-Walker models are solutions to EFE that are spatially isotropic and homogeneous.

multiverse—based on probabilistic judgments is warranted. But until then, it places such probabilistic arguments in cosmology at the forefront of both cosmology and debates in the philosophy of physics.

1.1.3 A Caveat about Cosmological Measures

Fine-tuning arguments in cosmology regarding the initial state of the universe or its global properties, as well as anthropic predictions, imply an appropriate probability measure over some parameter of interest. They imply that such a measure can exist, and that it can be well-defined and well-behaved. In all the papers in this dissertation, I take for granted that it makes sense to speak of a probability distribution over the space of solutions to EFE with respect to, e.g., the value of the cosmological constant (especially when I discuss anthropic arguments in Chapters 2, 3, and 4) or the set of values of different such parameters in a cosmological model (in paper 5).

There are, however, good reasons to doubt that such an assumption is tenable. Schiffrin and Wald (2012); Curiel (2015) assessed the technical and conceptual difficulties of defining the kind of measure used in such cosmological arguments. In particular, the families of spacetimes we work with and with which we carry out probabilistic reasoning are spaces of Lorentzian metrics on differential manifolds, which tend to be infinite-dimensional. However, "any translation-invariant measure on any reasonably well-behaved infinite-dimensional space assigns infinite measure to all open sets, unless the measure is the trivial measure," i.e., the one that assigns measure zero to every measurable set. (Curiel, 2015, § 3)

However, even if we restrict our family of models to finite-dimensional FLRW spacetimes with a scalar field, as with, e.g., the Gibbons-Hawking-Stewart measure, μ_{GHS},[3] serious difficulties remain as to the physical interpretation of such measures. As Schiffrin and Wald (2012) recalled, a measure like μ_{GHS} is constructed from a Hamiltonian for-

[3]See (Gibbons et al., 1987; Hawking and Page, 1988).

mulation of general relativity, which in statistical and Hamiltonian mechanics usually provides a canonical (Liouville) measure. But the requirements for assigning a Liouville measure to such a space, Schiffrin and Wald argued, face important difficulties. Even in this restricted space of solutions to EFE, the total measure of phase space is infinite, because the phase space of general relativity is noncompact. As a consequence, any proper subset of phase space is of measure (or probability) zero. One could use "regularization procedures" that roughly work as follows (see Schiffrin and Wald, 2012, § IV): if we want to assign a probability $p(X)$ to a property X within the space Γ, and if both X and $\Gamma \setminus X$ are of infinite measure, a regularization procedure approximates $p(X)$ by a nested sequence of finite-measure subsets $\{\Gamma_n\}$ with $\cup_n \Gamma_n = \Gamma$, and by defining $p(X)$ as $p(X) = \lim_{n\to\infty} \frac{\mu_{\text{GHS}}(X \cap \Gamma_n)}{\mu_{\text{GHS}}(\Gamma_n)}$. However, the result will depend on what is, arguably, an arbitrary choice of partition of Γ.[4]

However, even if there is no ambiguous way to define a probability based on the GHS measure, the physical significance of that measure is itself problematic. Indeed, as Schiffrin and Wald argued, arguments for a Liouville measure based on the dynamical evolution of the system of interest (here, the universe) do not apply when: "the system is not ergodic, (...) one has not waited a time much greater than the equilibrium time after the system was prepared," or when "the system has a time-dependent Hamiltonian that is varying on a time scale that is small or comparable to the equilibration time." (Schiffrin and Wald, 2012, § III) But in cosmology none of these conditions obtains, which precludes statistical equilibration.

1.1.4 Broader Epistemological Issues

The remarks we just saw in § 1.1.3 constitute a serious challenge to a legitimate and physically meaningful use of measures in cosmology (at least when those are defined over

[4]Schiffrin and Wald (2012) give as an example (Gibbons and Turok, 2008) and (Carroll and Tam, 2010), which gave very different results for the probability that the universe would have undergone a large number of e-foldings of inflation, based on two plausible constructions of $\{\Gamma_n\}$.

the space of solutions to EFE). As a consequence, they question the validity of many of the probabilistic arguments the papers in this dissertation touch on. Nevertheless, the relevance of this dissertation's conclusions is rooted in broader discussions. But, like the issue of the lawlikeness of cosmological laws previously illustrated, cosmology will elicit concept-clarification more forcefully than many other disciplines. Indeed, in cosmology,

> [p]resuppositions (regarding causality or time-order, for example) are made explicit; implications (such as the rejection of simultaneity at a distance) are explored. Consistency is tested. Relations between the most general principles of the theory and the conceptual framework within which the basic observations have to be situated, are worked out. (...) [T]his sort of conceptual analysis is akin to that which elsewhere defines the work of philosophy. (McMullin, 1981, 182)

Thus I hope to show in this dissertation that not only can philosophy be relevant to cosmology, but also cosmology to philosophy. Indeed, all these papers are concerned with the question of justifying probabilistic arguments based on pre-empirical notions (such as indifference, typicality, or simplicity), which are of broad philosophical interest and whose scope goes beyond cosmology.

The first two papers (i.e., Chapters 2 and 3) deal with the question of how to represent our credences, and, in particular, how to represent ignorance and indifference (see also below §§ 1.2.1, 1.2.2). A common criticism against a subjective interpretation of probability (i.e., the view that probabilities should be construed as representing an agent's degree of belief) claims that it is psychologically unrealistic to assign to our credence a precise numerical value (see, e.g., Kyburg, 1978). Moreover, for those who conceive of probabilities in terms of betting behavior, it would be more realistic to deal with an interval of betting prices (bounded by a selling price and a buying price), rather than a unique value (see Smith, 1961).

A bigger problem with the demand that our credences be represented by a precise value of a probability distribution is that this artificially precise value will artificially support unwarranted conclusions. This becomes clear when we want to represent our complete ignorance or indifference about, e.g., the likely value of some physical parameter. A common probabilistic interpretation of the principle of indifference recommends that we assign the same precise probability value to each event about which we are equally indifferent. In Chapter 3, I will discuss the contentious Doomsday argument, an argument yielding a prediction about the end date for humanity based only on the knowledge of how long it has existed and the assumption that we are typical members of this reference class (see below § 1.2.2). Similar arguments lead to notoriously arbitrary conclusions: if for instance, I am told that a factory makes dice whose side length is equally likely to be anywhere between 1cm and 2cm, I would not obtain the same probability distribution about what this factory produces if I am indifferent about these dice's *side length* or if I am indifferent about these dice's *volume* (van Fraassen, 1989). Thus one can see that it is possible, with this probabilistic interpretation of the principle of indifference, to obtain, as Fisher (1922, 325) wrote, "a vitally important piece of knowledge (...) out of complete ignorance" and arbitrary choices. For that reason, some have argued that a *more accurate* representation of our credences should be, in fact, *less precise*:

> As sophisticated Bayesians like Isaac Levi (1980), Richard Jeffrey (1983), Mark Kaplan (1998), have long recognized, the proper response to symmetrically ambiguous or incomplete evidence is not to assign probabilities symmetrically, but to refrain from assigning precise probabilities at all. (...) Imprecise credences have a clear epistemological motivation: they are the proper response to unspecific evidence" (Joyce, 2005, 171).[5]

Objections to imprecise probabilities assert that they contradict general principles of conditionalization. Indeed, in some circumstances, updating our credence after obtaining

[5]See also, e.g., (Levi, 1974; Walley, 1991; Joyce, 2010; Augustin et al., 2014).

new information results, somewhat counter-intuitively, in a less precise posterior credence; this is known as the "problem of dilation".[6] This issue has occupied a large part of the discussion about imprecise probabilities, and it led some to wonder, for instance, whether this reveals a more general problem with Bayesian updating, or if we ought to restrict the domain of reasonable credal states to those that preclude dilation (see Bradley, 2015, § 3.1 for a survey). But some contend that dilation is in fact a problem for conditionalization, or that it is even a bug of imprecise Bayesianism; they simply argue that "[t]he dilation-vulnerable parts of your prior conditional beliefs simply indicate cases where there is evidence of unknown value to be learned."[7] (Bradley and Steele, 2014b, 1301–2) Another objection to imprecise probabilities centers around possible, problematic consequences in decision making with imprecise probabilities. Adam Elga (2010), for instance, proposed a sequence of bets which, under certain circumstances, cannot result in the greatest possible gain, and which can even lead to a sure loss, if an agent's credences are allowed to be imprecise.[8] It is debatable that Elga's is in fact an objection strong enough to support the claim that credences *should* be precise. It has been argued indeed, e.g., in (Sahlin and Weirich, 2014), that there *are* decision making rules for imprecise probabilities that don't result in a sure loss.[9] Moreover, as argued in (Bradley and Steele, 2014a), even if there weren't such decision making rules, it would not be reasonable to reject imprecise probabilities as a whole on the grounds that they make some bad decisions merely *permissible*.

[6] Roger White (2010), for instance, made the claim that dilation violates the reflection principle, according to which an agent ought to have now a certain credence in a given proposition if she is certain she will have it at a later time (see van Fraassen, 1984). Jim Joyce, however, argued that this claim is incorrect: "Reflection does *not* tell you to have imprecise beliefs now if you know that you will have imprecise beliefs in the future. Rather, it says that your current credence for [a given hypothesis] should coincide with your current *expectation* of your future credence for [that hypothesis]." (Joyce, 2010, 303–304)

[7] This is arguably a deflationary view of the problem of dilation, but, according to these authors, it is the consensus view in statistics.

[8] In other words, he offered a variant of a diachronic "dutch book" argument.

[9] Namely, maximizing minimum expected utility. I will briefly discuss several possible decision making rules with imprecise probabilities in Chapters 2 and 3.

I will show in Chapter 3, on the other hand, that the conclusion of the Doomsday argument *cannot* be avoided by any Bayesian approach that represents credal states by single credence functions. I will show how imprecise probabilities allows one to avoid getting, as Fisher wrote, "a vitally important piece of knowledge (...) out of complete ignorance" (already cited above). Therefore, pace Elga, I will argue that, under certain circumstances, not only that our credence *should* be imprecise, but also, more specifically, that it should be represented by a set of probability distributions.

In addition to addressing the question of how we should represent credences, the papers in this dissertation raise the issue of whether ignorance or indifference should drive our inferences and predictions. In the arguments I will discuss, indifference claims can be made about physical events, but they can also be made about ourselves, viz., about our location as observers. Assumptions about the typicality of our location as observers made in cosmology (see §§ 1.2.3, 4) are sometimes taken as examples to assert the relevance of self-locating beliefs (see, e.g., Titelbaum, 2013). Such claims are part of a large literature that centers around the Sleeping Beauty problem (Elga, 2000; Lewis, 2001; Titelbaum, 2008) or other such thought experiments (such as in Elga, 2004), and which falls within a larger discussion on how to handle indexicals.[10] According to Ofra Magidor, much of the works on this topic share the view—which she calls "the myth of the *de se*"—that indexical propositions and, among them, self-locating propositions, require a special handling in epistemology and in confirmation:

> There is a special class of propositional (or "propositional-like") attitudes. These are self-locating or *de se* attitudes, ones that are typically expressed using indexical expressions such as "I" and "now" (...). Moreover, such attitudes pose a special challenge for our account of propositional attitudes. In other words, assume one starts with what might otherwise be considered an adequate account for standard (non *de se*) attitudes. Once we take on

[10] See, e.g., (Lewis, 1979) or more recently (Moss, 2012).

> board *de se* attitudes, this account ought to be fundamentally amended (. . .). (Magidor, 2014, 1)

Such calls to modify our non-indexical propositional attitudes by taking into account indexical elements were made explicitly in the context of cosmology. Nick Bostrom, most notably, argued that, because cosmological theories imply that, in a spatially infinite universe, any possible observation will almost certainly be made, they cannot "have *any* observational consequence *at all*" (Bostrom, 2002b, 608) unless we adopt "a methodology for evidence with a *de se* component." (op. cit., 621). Here is his claim in a nutshell:

> we must be careful about how we construe the evidence. We know not only that such-and-such observations are made (which I shall show is impotent as a basis for evaluating Big World theories): we also know that such-and-such observations are made *by us*. This indexical de se component of our evidence turns out to be crucial to cosmology, and recognizing this is the first step to the solution I shall propose.
>
> The second step is to formulate a new methodological principle that describes the probabilistic evidential bearing of (partly) indexical information on non-indexical hypotheses. (Bostrom, 2002b, 608–609, original emphasis)

Such claims imply that there are two, presumably distinct kinds of uncertainty: uncertainty "about what the world is like," and uncertainty "about one's own spatial or temporal location in the world" (Elga, 2000, 143). These two kinds of uncertainty are about what Magidor called non-indexical and, respectivelly, indexical propositional attitudes. Adam Elga (2000) formulated the Sleeping Beauty problem in order to show why our account of non-indexical propositional attitudes needs to be amended; he then argued for assumptions and methods specifically adapted to handle with self-locating uncertainty in order to solve that problem (introduced below in §§ 1.2.3, 4.2.1). If Elga and Bostrom are right, then solutions to the Sleeping Beauty problem (and, more generally, a methodology with a de se component) can help to place a priori constraints on our credence about the physical world, which could be especially relevant in cosmology.

Relatedly, invoking simplicity in theory choice and model selection aims at providing us with another a priori guiding principle. As we will see below in §§ 1.2.4, 5, this epistemological inclination is sometimes justified—at least in the science literature—by an appeal to Ockham's razor. Simplicity is usually considered to be "a standard criterion for evaluating the adequacy of a theory" (Kuhn, 1977, 357) or for model selection, but as Kuhn argued, theoretical simplicity can take on different meanings: a theory can afford us more easily computable problem-solving methods, it can constitute a more unifying explanation to more phenomena, or it can be more ontologically parsimonious. As Kuhn argued,

> [s]implicity, however, favored Copernicus, but only when evaluated in a quite special way. If, on the one hand, the two systems were compared in terms of the actual computational labor required to predict the position of a planet at a particular time, then they proved substantially equivalent. Such computations were what astronomers did, and Copernicus's system offered them no labor-saving techniques; in that sense it was not simpler than Ptolemy's. If, on the other hand, one asked about the amount of mathematical apparatus required to explain, not the detailed quantitative motions of the planets, but merely their gross qualitative features—limited elongation, retrograde motion and the like—then, as every schoolchild knows, Copernicus required only one circle per planet, Ptolemy two. In that sense the Copernican theory was the simpler, a fact vitally important to the choices made by both Kepler and Galileo and thus essential to the ultimate triumph of Copernicanism. But that sense of simplicity was not the only one available, nor even the one most natural to professional astronomers, men [*sic*] whose task was the actual computation of planetary position. (Kuhn, 1977, 358)

Yet according to the same Kuhn, "Copernicus, too, was forced to use minor epicycles and eccentrics. His full system was little if any less cumbersome than Ptolemy's had been. Both employed over thirty circles; there was little to choose between them in economy." (Kuhn, 1959, 169) One can then wonder what sense of simplicity we ought to adopt, and on what grounds.

In this dissertation, I will give a generally critical assessment on the issue of simplicity as well as about other pre-empirical notions and their role in inference and confirmation. I discuss, particularly in Chapter 2 (see also § 1.2.1), claims that are very critical of probabilistic reasoning in general, in part because, as we saw, the formal requirements of a probability distribution—such as additivity—place too strong a constraint on the representation of our credences. One might draw similarly critical conclusions from the limitations of statistical and Bayesian analysis in model selection which I point out in Chapter 5 (see also § 1.2.4). However, in these two papers, while I will acknowledge limitations and inadequacies in the use of probabilities in induction, I will show how this same inductive framework offers us a way out of these difficulties, and what legitimate role probabilistic reasoning can play in cosmology.

1.2 Outline of the Thesis

Chapters one, two, and three will deal with the intersection of two distinct problems: the problem of accounting for selection effects and the problem of representing ignorance or indifference in probabilistic inferences. These two problems meet in different places in the cosmology literature. A first case is that of anthropic reasoning used to predict cosmological parameters such as the cosmological constant. In the absence of fundamental theories to explain, e.g., the value of the vacuum energy density ρ_V, it has been argued that we should conditionalize the probability of different values of ρ_V on the number of observers they allow (Weinberg, 1987). Without a well-motivated measure for the cosmological constant, such arguments appeal to a uniform prior probability distribution in order to represent our lack of knowledge and in order to obtain a prediction. The prediction thus obtained, however, reflects the inadequacy of uniform probability distributions as representations of indifference. John Norton (2010) has argued that this inadequate uniform probability distribution, which turns indifference into improbability, is a fundamental flaw of probabilistic inference or in general any inductive framework using additive measures. Norton's criticism is a serious challenge to Bayesianism. In chapter one, I will examine how imprecise probabilities (whereby credences are represented by a family of probability distributions) fare as an inductive framework and avoids such unwarranted inferences. In particular, I will ask whether they can meet Norton's criteria for a candidate for representation of neutral support. In chapter two, I will detail how this framework allows us to successfully address Doomsday arguments in a way no Bayesian approach that represents credal states by single credence functions could.

Questions about the role of our place as observers arise in other parts of cosmology. We may be uncertain about how representative or typical our observations are, depending on whether we find ourselves in a typical region of our universe. We may also be uncertain about our location as observers if we know that there may be copies of ourselves, i.e., other observers having experiences identical to our own. Or we may also have an uncertainty about our location as observers if in the multiverse different "universes" (or "pocket universes") are possible and which would be compatible with the data we have. Such worries are sometimes discussed as questions about self-locating uncertainty, handled by endorsing a kind of Copernican Principle, i.e., by simply assuming that we are "typical observers". Some arguments used to justify such assumptions closely resemble solutions to puzzles of self-locating beliefs well known to philosophers, such as the Sleeping Beauty problem. Chapter three will investigate whether and to what extent such self-locating uncertainty is distinct from our uncertainty about what the world is like in such a way that these two kinds of uncertainty require different handling in confirmation theory.

The problems raised in the first three chapters join more general criticisms about the role of probabilistic inference in cosmology. A more general issue concerns the inadequacy of Bayesian model selection criteria in the absence of well-motivated measures for different cosmological models. Bayesian model selection extends inferences about parameter estimation to the space of models (Trotta, 2012). Information criteria commonly used for model selection (such as the Akaike Information Criterion or the Bayesian Information Criterion) tend to optimize the number of free parameters. But indiscriminate applications of what is taken to be Ockham's razor can select physically implausible models (Efstathiou, 2008). The last chapter will examine the possibility for Bayesian model selection to circumvent the lack of well-motivated priors. In particular, I will evaluate to what extent the virtue of unification as a criterion for model selection—defended by Myrvold (2003)—overcomes the limitations of usual Bayesian information criteria.

1.2.1 The Bayesian Who Knew Too Much

Ongoing work by cosmologists (see e.g., Carr, 2007) purports to explain an allegedly surprising coincidence—the fact that physics has produced a universe capable of hosting life. Indeed, if the value of some physical constants, left indeterminate by our current theories, were only slightly different, life wouldn't be possible. These authors claim that this state of affairs corresponds to a low-probability fine-tuning of cosmic parameters. In such works, and apparently in accordance with Bayesian analysis, a theory that increases the odds of what these authors consider to be a surprising cosmic arrangement should be favored over those which don't. The initial low probability values that motivate the need for an explanation, however, is meant to represent our ignorance or indifference about what value of a parameter we should expect.

Steven Weinberg (1987) resorted to a similar kind of confirmation when he appealed to anthropic constraints (i.e., constraints related to the possibility for life to exist) in order to explain the value of the cosmological constant (i.e., the vacuum energy density

ρ_V), left indeterminate by existing theories. Anthropic considerations provide bounds on ρ_V, which can be neither too large (because then galaxies couldn't form) nor too small (because then observers wouldn't have time to arise before the universe recollapses). However, the anthropically-allowed range is quite large, and, in the absence of further theories, it leaves us with many equipossible values for ρ_V. Furthermore, by conditionalizing on the number of observers each value of ρ_V allows for, the initial uniform distribution is turned into a prediction. Later observations of the cosmological constant have been taken to vindicate this reasoning (see, e.g., Weinberg, 2007).

However, it is debatable whether a precise probability distribution over admissible values of some physical parameters or their ratios can reasonably reflect a reasonable prior credal state. Moreover, it is questionable that the conditions for life (e.g., the particular value of certain physical constants) should be assigned a *low* rather than a *neutral* probability value, if these parameters are left indeterminate by our theories. John Norton has argued that Bayesianism cannot handle ignorance adequately due to its inability to distinguish between neutral and disconfirming evidence. He argued that, in anthropic reasoning as in other cases, this inability allows one to draw unwarranted conclusions from a *lack of knowledge*. Because it stems from the requirement of additivity, it is, on Norton's account, the sign of a fundamental flaw of Bayesian calculus. In several works, Norton (2007, 2008, 2010) has introduced criteria for a candidate for representation of neutral evidential support.

Joyce (2010) defended a version of Bayesianism without sharp credences. An imprecise probability—or a 'representor' following (van Fraassen, 1990)—consisting of a family of credal functions can be thought of as representing the credal state of a single agent composed of several members who don't necessarily agree with each other (e.g., a jury or a committee). Representing a credal state by means of a family of credal functions allows one to give a probabilistic representation of ignorance that 1) distinguishes stochastic independence from unknown interaction between parameters, and 2) better models the non-informativeness of priors corresponding to a credal state of ignorance or indifference. This analysis allows us to see that, contrary to Norton's claim, it isn't necessarily Bayesian reasoning, but rather an inadequate application of it that may be at the origin of confusions in such instances of anthropic reasoning.

We can see to what extent imprecise probabilities fulfill Norton's proposal, and whether or not the latter should be amended. In particular, we can see that, if adopted, the imprecise model generally agrees with Norton's representation of neutral support but demands that his criterion of self-duality be reformulated. The dual of a measure of belief is a measure of disbelief, and for Norton a representation of indifference or ignorance must be self-dual. Because of the requirement of additivity, no measure can be self-dual. Applied to the imprecise model, a credal set C representing ignorance is self-dual if it contains probability functions c_i and their duals c_i' such that $\forall c_i \in C, c_i'$ is such that $\forall \alpha, c_i(\alpha) = c_i'(\neg\alpha)$ and $c_i(\neg\alpha) = c_i'(\alpha)$. It is clear that we cannot have such a set since the dual of a probability function cannot be a probability function. But I will argue that we need not impose such a strong constraint on credal sets.

Depending on what decision-making criterion with imprecise probabilities one chooses, it is possible to construct a representation of indifference by means of credal sets that meet all or almost all of Norton's criteria, and it can do so without compromising Bayesianism altogether. It only requires that we demand neither that credences be sharp nor that a unique representation be applicable to all cases of ignorance or indifference (i.e., that self-duality be abandoned).

1.2.2 Blurring Out Cosmic Puzzles

The previous section briefly introduced anthropic reasoning as an example of probabilistic reasoning that allows one to draw unwarranted conclusions from a lack of knowledge. The Doomsday argument is another example of such puzzles of probabilistic confirmation. The Doomsday argument is in fact a family of arguments given to help us determine the end date for humanity based only on the knowledge of the time elapsed since humanity's advent. Both puzzles make 'cosmic' predictions based on typicality assumptions: by conditionalizing on the number of observers allowed for by a given value of the cosmological constant, anthropic reasoning assumes that we are typical observers; Doomsday-type arguments assume that our observation time is taken at random among all of humanity's. Thus, the Doomsday argument and anthropic reasoning share a similar structure: 1) a uniform prior probability distribution reflects an initial state of ignorance or indifference, and 2) an appeal to typicality or mediocrity is used to make a prediction. This is puzzling: these two assumptions of indifference and typicality are meant to express neutrality, and yet from them alone we seem to be getting a lot of information. But assuming neutrality *alone* shouldn't allow us to learn anything! One way of formulating both of these arguments (anthropic reasoning and the Doomsday argument) makes them a straightforward application of Bayesianism, yet the conclusion is so absurd as to cast doubt on Bayesianism itself.

Much of the philosophical discussion around these arguments has focused on the validity of the typicality assumption. For instance, Nick Bostrom (2002a) offered a challenge to what he calls the Self-Sampling Assumption (SSA), according to which "one should reason as if one were a random sample from the set of all observers in one's reference class." In order to avoid the consequence of the Doomsday argument, Bostrom suggested to adopt what he calls the Self-Indicating Assumption (SIA): "Given the fact that you exist, you should (other things equal) favor hypotheses according to which many observers exist over hypotheses on which few observers exist." But this arguably ad hoc solution leads to the unpalatable consequence that a preference for a cosmological scenario over another can be established entirely a priori. Relatedly, Radford Neal (2006) argued that conditionalizing on non-indexical information (i.e., all the information at the disposal of the agent formulating the argument, including all their memories), reproduces the effects of Bostrom's SIA without its adhocness.

However, these solutions to an assumption of typicality cannot satisfactorily avoid the conclusions of all versions of the Doomsday argument. A version put forth by Richard Gott (1994) which starts from an assumption about our birth rank (instead of that of our position as observers) among all of humankind will yield a prediction for humanity's end date whatever assumption we make as to our typicality or atypicality.

Imprecise probabilities allow us to better represent our initial state of ignorance or indifference, with a credal set of functions that disagree with each other. If the *only* thing we want our prior credal state to represent is that we know neither whether nor when humanity will end, then we ought to include in our prior set an *infinity* of normalizable credal functions about the possible total number of humans. In order to be normalizable, all these functions have to tend to zero as the total number of possible humans goes to infinity. At any confidence level, not all functions in that set will agree on an upper bound on that total number. This result, however, would be impossible to obtain with a single prior probability distribution, with which we would have no choice but to predict an end date for humankind. Relatedly, in a bounded case (e.g., anthropic predictions for values of the cosmological constant), it is possible to construct an imprecise prior credal set that results in a very weak—or even no—confirmation of anthropic hypotheses.

I will show that imprecise probabilities can dissolve these puzzles better than precise probabilities could. Philosophical discussions about the value of the imprecise model usually center around the difficulty to define updating rules that don't contradict general principles of conditionalization. But the expressive richness of this framework and its ability to solve such puzzles of confirmation and avoid unwarranted conclusions should be considered as a crucial feature of the imprecise model and count in its favor.

1.2.3 Lost in the Multiverse: Self-locating uncertainty, typicality, and observation bias

There are in the cosmology literature several kinds of arguments referring to self-locating uncertainty. In the multiverse, different "universes" are possible: different "pocket-universes" may have different fundamental physical parameters. We don't know how to locate ourselves in this ensemble of possible universes. We don't know how to locate ourselves within our own universe either. We don't know if we are typical observers and if we can safely assume that the physical laws we draw from our observations hold elsewhere or if, on the contrary, our view on the world is particularly biased spatially or temporally.

A kind of self-locating uncertainty that is sometimes invoked in such a discussion comes from the possible existence of *copies of myself*. For instance, if the multiverse theory is true, then there are other pocket universes where reside observers having experiences indistinguishable from my own. As a consequence, we should be uncertain as to which of these pocket universes we inhabit. If one of these pocket universes is more favorable to the advent of life than the other (because of a difference in the value of some physical parameter), then this pocket universe is more likely to host a greater number of such copies of myself. This adds to the uncertainty about my location: I don't know which of these pocket universes I may find myself in, and, within this pocket universe, I don't know which of the copies of myself I am.

This kind of uncertainty is used for instance in anthropic reasoning, where predictions are obtained by conditionalizing a prior probability distribution for a given cosmological parameter on the number of observers it allows for (see, e.g., Weinberg, 1987; Bousso, 2006). Moreover, according to, e.g., Max Tegmark, there is even a real possibility that, in a universe that is large enough, "your closest identical copy is 10 to the 10^{28} meters away," (Tegmark, 2003, 48); i.e., within that distance, each of us has a doppelgänger with *exactly* the same experiences and memories.

There are two main issues concerning all the above questions and arguments that require clarification. One is the status of self-locating beliefs and their role in induction and confirmation. The first question I will address is how distinct are the following "two sorts of uncertainty": one "about what the world is like" and another one "about one's own spatial and or temporal location in the world" (Elga, 2000, 143). Should the realization that I *could* very possibly be somewhere else affect my beliefs about the world? That is, does the fact that a physical model is more favorable to the advent of observers having experiences indistinguishable from my own, or that it generates more observing standpoints resulting in an experience identical to my own, affect my credence in that model? And if so, does that mean that my self-locating uncertainty should have a role in confirmation and induction distinct from other kinds of uncertainty?

A second, related question concerns *how* to handle self-locating uncertainty. In particular, I will ask whether and when assumptions of typicality (or a Copernican principle, or an assumption of mediocrity) are warranted or even required to characterize our self-locating uncertainty.

I will claim that many of these issues that are often taken to be about "self-locating uncertainty" are in fact problems about how to handle observation bias or about uncertainty about what the world is like. Consequently they don't constitute a new or distinctive challenge or source of rules for confirmation theory.

These arguments involving our "location" broadly construed are not particularly new, whether in cosmology or in philosophy.[11] But the topic of self-locating beliefs (or, following Lewis (1979), "*de se* beliefs") and whether they bear on *de re* beliefs has recently generated much discussion in philosophy. Arguments given as solutions to the Sleeping Beauty problem—a thought experiment introduced as an illustration of the distinctness of self-locating uncertainty and its bearing on our credence about the world—are still

[11]For a critique of the anthropic principle, see, e.g., (Earman, 1987), for one of the Copernican principle, see, e.g., (Beisbart, 2010).

debated, in philosophy and cosmology alike.

The Sleeping Beauty problem is sometimes explicitly invoked in the cosmology literature to justify the assumption—made by, e.g., Tegmark (2003); Bousso et al. (2008); Page (2010); Freivogel (2011)—that we are equally likely in the multiverse to be anywhere consistent with our data. Beauty will be put to sleep for three days on Sunday night. Right after she falls asleep, a fair coin will be tossed to determine how many times she will be briefly woken. If the coin toss results in *Heads*, Beauty will be briefly woken only once, on Monday. If *Tails*, she will be woken twice: once on Monday, and once on Tuesday. But after each waking, Beauty will be put back to sleep with a drug that makes her completely forget about that waking. Now, right after she has been woken but without having been told what day it is, what should Beauty's credence be that the coin came up *Heads*? According to Elga (2000), she should have a credence of 1/3 that the result of the coin toss was *Heads*. I will show that appeals to this solution in the cosmology literature as a justification for the claim that all three possible awakenings are *a priori* equally likely, rest on confusions about the Sleeping Beauty problem and about Elga's argument. I will show how this puzzle clarifies the nature and role of self-locating beliefs, and in particular how typicality assumptions in cosmology come down to asserting a "mere personal preference for theories in which we are typical of something." Hartle and Srednicki (2007, 1)

In order to assess what role typicality assumptions may play in cosmological model selection within our "pocket universe," and when they might be warranted, I will examine arguments to defend the Copernican principle. Cosmological models as well as the interpretation of our cosmological data usually rest on the adoption of the Copernican principle, i.e., on the assumption that the apparent spatial homogeneity and isotropy doesn't stem from our having a very special point of view on our cosmic neighborhood. This has very real implications for our choice of cosmological models. Indeed, if the Copernican principle holds, then our observations lead to the conclusion that the expansion of the observable universe is accelerating, and thus they support the existence of dark energy. However, assuming that we are at the center of a large 'cosmic bubble' could explain the same observations that are taken to support dark energy. The adoption of the Copernican principle, however, forces us to discard this alternative hypothesis, and to posit this mysterious dark energy instead. According to arguments about how self-locating beliefs should bear on our knowledge of what the world is like, the fact that the number of locations we could find ourselves in is much larger if we adopt the Copernican principle than in the "cosmic bubble" scenario should favor the former hypothesis over the latter. I will show that it is neither solutions to the Sleeping Beauty problem nor in general considerations about self-locating uncertainty, but rather more evidence (see, e.g., Uzan et al., 2008) that will help us adopt or reject the Copernican principle.

1.2.4 Simplicity and Unification in Cosmological Model Selection

The use of Bayesian methods in cosmology is a relatively new field of research (see, e.g., Trotta, 2008, 2012; Liddle, 2009; Hobson et al., 2010, for recent reviews). In the absence of well-motivated prior probabilities for cosmological models, Bayesian and statistical model selection appeals to, e.g., the "Astronomer's prior," a uniform distribution over a given range of models. Because selecting cosmological models based only on how likely they make our data is almost guaranteed to favor models that overfit the data, cosmologists rely on a sense of "simplicity" (or "Ockham's razor") to weigh a model's accuracy against its number of free parameters. For that same reason, they sometimes use statistical tools—such as the Bayesian Information Criterion (BIC) or the Akaike Information Criterion (AIC)—as an Ockham's razor.

However, although these methods purport to minimize our reliance on a choice of prior, they are not able to isolate physically implausible models. As a result, in cosmology where well-motivated priors or statistical data are lacking, such model selection criteria have been said to be no more than prior selection, casting doubt on the possibility to have meaningful Bayesian cosmological model selection altogether (see, e.g., Efstathiou, 2008; Linder and Miquel, 2008).

I will first assess the claim that what such Bayesian and statistical methods do is indeed to seek a balance between a model's fit-to-data and its complexity (defined in terms of the number of its free parameters). We will see that the notion of "simplicity" is ambiguous and, as argued by Norton (2012), a surrogate for background information. In addition, we will see that, strictly speaking, it is not simplicity that these model selection criteria are after.

Regardless of the question of simplicity, we will see that statistical and Bayesian model selection faces serious challenges, especially in cosmology, due to the limited data sample available to us and the lack of theoretical background. Therefore, I will consider what other Bayesian model selection method might fare better in cosmology.

In contrast with model selection driven by the search for parsimony in the number of free parameters, Wayne Myrvold (2003) has given a Bayesian account of a model's ability to unify different sorts of phenomena, regardless of the number of parameters. The measure of information used in this account is between a set of phenomena and another, otherwise unrelated set of phenomena, given a certain model. This measure gives a formal character to the idea of consilience of inductions (Whewell, 1847), according to which "a consilience of inductions would occur when the values of certain parameters can be determined from two different sorts of phenomena, and the values determined from one class of phenomena agree with those determined from another." (Myrvold, 2003, 418) Parameter estimation plays a role in assessing the unifying power of a model, but on this account it is the improved relationship between parameters of different kinds that will provide support to a choice of model. I wish to explore if this Bayesian criterion of unification can be better suited to some probabilistic inferences in cosmology.

Bibliography

Albert, D. Z. (2000). *Time and Chance*. Harvard University Press.

Augustin, T., Coolen, F. P., de Cooman, G., and Troffaes, M. C., editors (2014). *Introduction to Imprecise Probabilities*. Wiley & Sons.

Beisbart, C. (2010). Can We Justifiably Assume the Cosmological Principle in Order to Break Model Underdetermination in Cosmology? *Journal for General Philosophy of Science*, 40(2):175–205.

Bostrom, N. (2002a). *Anthropic Bias: Observation Selection Effects in Science and Philosophy*. Routledge.

Bostrom, N. (2002b). Self-Locating Belief in Big Worlds: Cosmology's Missing Link to Observation. *The Journal of Philosophy*, 99(12):607–623.

Bousso, R. (2006). Holographic Probabilities in Eternal Inflation. *Physical Review Letters*, 97:191302 1–4.

Bousso, R., Freivogel, B., and Yang, I.-S. (2008). Boltzmann Babies in the Proper Time Measure. *Physical Review D*, 77(10):103514.

Bradley, S. (2015). Imprecise Probabilities. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Summer2015 edition.

Bradley, S. and Steele, K. (2014a). Should Subjective Probabilities Be Sharp? *Episteme*, 11(3):277–289.

Bradley, S. and Steele, K. (2014b). Uncertainty, Learning, and the "Problem" of Dilation. *Erkenntnis*, 79(6):1287–1303.

Carr, B. (2007). The Anthropic Principle Revisited. In Carr, B., editor, *Universe or Multiverse?*, chapter 5, pages 77–89. Cambridge University Press.

Carroll, S. M. and Tam, H. (2010). Unitary Evolution and Cosmological Fine-Tuning. *arXiv preprint arXiv:1007.1417*, pages 1–34.

Curiel, E. (2015). Measure, Topology and Probabilistic Reasoning in Cosmology (unpublished manuscript). (April 22).

Disney, M. (2000). The Case Against Cosmology. *General Relativity and Gravitation*, 32(6):1125–1134.

Earman, J. (1987). The SAP Also Rises: A Critical Examination of the Anthropic Principle. *American Philosophical Quarterly*, 24(4):307–317.

Efstathiou, G. (2008). Limitations of Bayesian Evidence applied to cosmology. *Monthly Notices of the Royal Astronomical Society*, 388:1314–1320.

Elga, A. (2000). Self-Locating Belief and the Sleeping Beauty Problem. *Analysis*, 60(2):143–147.

Elga, A. (2004). Defeating Dr. Evil with Self-Locating Belief. *Philosophy and Phenomenological Research*, LXIX(2):383–396.

Elga, A. (2010). Subjective Probabilities Should Be Sharp. *Philosopher's Imprint*, 10(5).

Ellis, G. F. R. (2007). Issues in the Philosophy of Cosmology. In Butterfield, J. and Earman, J., editors, *Handbook of the Philosophy of Science, Philosophy of Physics, Part B*, pages 1183–1283. Elsevier, Amsterdam.

Feynman, R. P. (1967). *The Character of Physical Law*. MIT Press, Cambridge, MA.

Fisher, R. A. (1922). On the Mathematical Foundations of Theoretical Statistics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 222:309–368.

Freivogel, B. (2011). Making Predictions in the Multiverse. *Classical and Quantum Gravity*, 28(204007):1–15.

Gibbons, G., Hawking, S. W., and Stewart, J. (1987). A Natural Measure on the Set of All Universes. *Nuclear Physics B*, 281:736–751.

Gibbons, G. and Turok, N. (2008). Measure Problem in Cosmology. *Physical Review D*, 77(6):063516.

Gott, J. R. (1994). Future Prospects Discussed. *Nature*, 368(March):108.

Guth, A. H. (1981). Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems. *Physical Review D*, 23(2):347–356.

Hartle, J. B. and Srednicki, M. (2007). Are We Typical? *Physical Review D*, 75(12):123523.

Hawking, S. W. and Page, D. N. (1988). How Probable Is Inflation? *Nuclear Physics B*, 298:789–809.

Hobson, M. P., Liddle, A. R., Mukherjee, P., and Parkinson, D., editors (2010). *Bayesian Methods in Cosmology*. Cambridge University Press, Cambridge.

Jeffrey, R. (1983). Bayesianism with a Human Face. In Earman, J., editor, *Testing Scientific Theories*, Minnesota Studies in the Philosophy of Science, Vol. 10, chapter 7, pages 133–156. University of Minnesota Press, Minneapolis.

Joyce, J. M. (2005). How Probabilities Reflect Evidence. *Philosophical Perspectives*, 19:153–178.

Joyce, J. M. (2010). A Defense of Imprecise Credences in Inference and Decision Making. *Philosophical Perspectives*, 24(1):281–323.

Kaplan, M. (1998). *Decision Theory as Philosophy*. Cambridge University Press, Cambridge.

Kuhn, T. S. (1959). *The Copernican Revolution*. Random House, New York, NY.

Kuhn, T. S. (1977). Objectivity, Value Judgment and Theory Choice. In *The Essential Tension: Selected Studies in the Scientific Tradition*, pages 356–367. University of Chicago Press, Chicago.

Kyburg, E. H. (1978). Subjective Probability: Criticisms, Reflections, and Problems. *Journal of Philosophical Logic*, 7(1):157–180.

Levi, I. (1974). On Indeterminate Probabilities. *The Journal of Philosophy*, 71(13):391–418.

Levi, I. (1980). *The Enterprise of Knowledge*. MIT Press, London.

Lewis, D. (1979). Attitudes De Dicto and De Se. *The Philosophical Review*, 88(4):513–543.

Lewis, D. (2001). Sleeping Beauty: Reply to Elga. *Analysis*, 61(3):171–176.

Liddle, A. R. (2009). Statistical Methods for Cosmological Parameter Selection and Estimation. *Annual Review of Nuclear and Particle Science*, 59:95–114.

Linder, E. and Miquel, R. (2008). Cosmological Model Selection: Statistics and Physics. *International Journal of Modern Physics D*, 17(12):2315–2324.

Magidor, O. (2014). The Myth of the De Se. *Philosophical Perspectives (forthcoming)*, pages 1–53.

McCoy, C. D. (2015). Does Inflation Solve the Hot Big Bang Model's Fine-Tuning Problems? *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 51(August):23–36.

McMullin, E. (1981). Is Philosophy Relevant To Cosmology? *American Philosophical Quarterly*, 18(3):177–189.

Moss, S. (2012). Updating as Communication. *Philosophy and Phenomenological Research*, LXXXV(2):225–248.

Munitz, M. K. (1952). Scientific Method in Cosmology. *Philosophy of Science*, 19(2):108–130.

Munitz, M. K. (1962). The Logic of Cosmology. *British Journal for the Philosophy of Science*, 13(49):34–50.

Myrvold, W. C. (2003). A Bayesian Account of the Virtue of Unification. *Philosophy of Science*, 70(2):399–423.

Neal, R. M. (2006). Puzzles of Anthropic Reasoning Resolved Using Full Non-indexical Conditioning. *arXiv:math/0608592*, pages 1–56.

Norton, J. D. (2007). Probability Disassembled. *The British Journal for the Philosophy of Science*, 58(2):141–171.

Norton, J. D. (2008). Ignorance and Indifference. *Philosophy of Science*, 75(1):45–68.

Norton, J. D. (2010). Cosmic Confusions: Not Supporting versus Supporting Not. *Philosophy of Science*, 77(4):501–523.

Norton, J. D. (2012). Simplicity as a Surrogate. In *The Material Theory of Induction (forthcoming)*, chapter 6, pages 1–46.

Page, D. N. (2010). Born's Rule Is Insufficient in a Large Universe. *arXiv preprint arXiv:1003.2419*, pages 1–7.

Penrose, R. (1979). Singularities and Time-Asymmetry. In Hawking, S. W. and Israel, W., editors, *General Relativity: An Einstein Centenary Survey*. Cambridge University Press.

Sahlin, N. E. and Weirich, P. (2014). Unsharp Sharpness. *Theoria (Sweden)*, 80(1):100–103.

Schiffrin, J. S. and Wald, R. M. (2012). Measure and Probability in Cosmology. *Physical Review D*, 86(2):023521.

Smeenk, C. (2008). The Logic of Cosmology Revisited (unpublished manuscript).

Smeenk, C. (2013). Philosophy of Cosmology. In Batterman, R., editor, *Oxford Handbook for the Philosophy of Physics*, chapter 17, pages 607–652. Oxford University Press, Oxford.

Smeenk, C. (2014). Predictability Crisis in Early Universe Cosmology. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 46(May):122–133.

Smith, C. A. (1961). Consistency in Statistical Inference and Decision. *Journal of the Royal Statistical Society. Series B (Methodological)*, 23(1):1–37.

Tegmark, M. (2003). Parallel Universes. *Scientific American*, 288(5):40–51.

Titelbaum, M. G. (2008). The Relevance of Self-Locating Beliefs. *Philosophical Review*, 117(4):555–606.

Titelbaum, M. G. (2013). Ten Reasons to Care about the Sleeping Beauty Problem. *Philosophy Compass*, 8(11):1003–1017.

Torretti, R. (2000). Spacetime Models for the World. *Studies In History and Philosophy of Science Part B:*, 31(2):171–186.

Trotta, R. (2008). Bayes in the Sky: Bayesian Inference and Model Selection in Cosmology. *Contemporary Physics*, 49(2):71–104.

Trotta, R. (2012). Cosmological Bayesian Model Selection: Recent Advances and Open Challenges. In Feigelson, E. D. and Babu, G. J., editors, *Statistical Challenges in Modern Astronomy V*, volume 902 of *Lecture Notes in Statistics*, pages 127–140. Springer New York, New York, NY.

Uzan, J.-P., Clarkson, C., and Ellis, G. F. R. (2008). Time Drift of Cosmological Redshifts as a Test of the Copernican Principle. *Physical Review Letters*, 100(19):191303–1–4.

van Fraassen, B. C. (1984). Belief and the Will. *The Journal of Philosophy*, 81(5):235–256.

van Fraassen, B. C. (1989). Indifference: The Symmetries of Probability. In *Laws and Symmetry*, chapter 12, pages 293–317. Oxford University Press.

van Fraassen, B. C. (1990). Figures in a Probability Landscape. In Dunn, J. and Gupta, A., editors, *Truth or Consequences: Essays in Honor of Nuel Belnap*, pages 345–356. Kluwer Academic Press, Amsterdam.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities.* Chapman and Hall, London.

Weinberg, S. (1987). Anthropic Bound on the Cosmological Constant. *Physical Review Letters*, 59(22):2607–2610.

Weinberg, S. (2007). Living in the Multiverse. In Carr, B., editor, *Universe or Multiverse?*, chapter 2, pages 29–42. Cambridge University Press.

Whewell, W. (1847). *The Philosophy of the Inductive Sciences, Founded upon Their History.* J.W. Parker, London.

White, R. (2010). Evidential Symmetry and Mushy Credence. *Oxford Studies in Epistemology*, 3:161–186.

Chapter 2

The Bayesian Who Knew Too Much[1]

2.1 Introduction: Bayesian Reasoning and Ignorance

If you ask me what the probability is of rolling a 2 with a throw of a fair, cubic die, I will answer, "$\frac{1}{6}$." Now imagine that we are having a conversation about Shakespeare's play *The Tempest* and you ask me to assign a probability to the claim that the play has 16,633 words. My response would be that I simply do not have a clue. My knowledge of Elizabethan drama would allow me to say that the value falls within some interval, in which all values seem equally plausible. However, I would not be able to assign a probability to the claim that the play has *exactly* 16,633 words. If you insist on a probability, I would respond—echoing David Albert—by asking, "What part of 'I don't have a clue' do you not understand?"

The simplest version of Bayesianism is ill-equipped to handle such a case. Reasoning with the principle of indifference, as in the case of the die, motivates a uniform probability distribution over some interval of values; if there are 20,000 equiprobable numbers in the plausible range of number of words for that play, each number will be assigned a probability of $\frac{1}{20{,}000}$. Thus according to usual Bayesian confirmation theory, learning the exact length of *The Tempest* is equivalent to confirming a claim with a very low probability. In contrast, suppose that we are discussing whether or not the 1956 movie

[1] This chapter was published in May 2015 in *Synthese 192* (5): 1527–1542. Section 2.4 below is an excerpt of § 3.3.

Forbidden Planet is a good adaptation of *The Tempest*, and you ask me to assign a probability to the claim that Shakespeare himself authored that movie's screenplay. I would *not* be indifferent about that claim; I would assign to it a very low probability, and I would be very surprised if it were confirmed. Yet orthodox Bayesianism does not allow one to distinguish these two confirmations: it sees them both as the confirmation of a low-probability claim, thereby justifying in both cases a same sense of surprise. One may legitimately be surprised by the confirmation of a low-probability proposition, but not by the confirmation of something about which we are entirely *indifferent*. In the absence of a representation of *neutral* degree of belief, we have no choice but to treat large numbers of alternatives about which we are equally *indifferent* as if they all were *improbable* propositions. Consequently, any hypothesis—however unlikely—that lends more support to the observed value than was initially given by the constant probability distribution will be seen as significantly confirmed.

Steven Weinberg (1987) resorted to a similar kind of confirmation when he appealed to anthropic constraints (i.e., constraints related to the possibility for life to exist) in order to explain the value of the cosmological constant (i.e., the vacuum energy density ρ_V), left indeterminate by existing theories. Anthropic considerations provide bounds on ρ_V, which can be neither too large (because then galaxies could not form) nor too small (because then observers would not have time to arise before the universe recollapses). However, the anthropically allowed range is quite large, and, in the absence of further theories, it leaves us with many equipossible values for ρ_V. Weinberg further argued that anthropic considerations may have a stronger, predictive role. The idea is that we should conditionalize the probability of different values of ρ_V on the number of observers they allow: the most likely value of ρ_V is the one that allows for the largest number of galaxies (taken as a proxy for the number of observers).[2] The probability measure for ρ_V is then

[2]This assumption is contentious (see, e.g., (Aguirre, 2001) for an alternative proposal).

as follows:

$$\mathrm{d}p(\rho_V) = \nu(\rho_V) \cdot p_\star(\rho_V)\,\mathrm{d}\rho_V,$$

where $p_\star(\rho)\,\mathrm{d}\rho_V$ is the prior probability distribution, and $\nu(\rho_V)$ the average number of galaxies which form for ρ_V. By assuming that there is no known reason why the likelihood of ρ_V should be special at the observed value, and because the allowed range of ρ_V is very far from what we would expect from available theories, Weinberg argued that it is reasonable to assume that the prior probability distribution is constant within the anthropically allowed range, so that $\mathrm{d}p(\rho_V)$ can be calculated as proportional to $\nu(\rho_V)\,\mathrm{d}\rho_V$ (Weinberg, 2000). Moreover, by assuming that we are typical observers (and thereby adopting what Alexander Vilenkin (1995) called a 'principle of mediocrity'), Weinberg predicted that the observed value of this parameter should be close to the mean of the anthropically allowed values. The initial uniform distribution is turned into a prediction, a sharply peaked distribution around a preferred value. Later observations of the cosmological constant have been taken to vindicate this reasoning (see, e.g., Weinberg, 2007).[3]

John Norton (2010) criticized the probabilistic representation of ignorance on which such arguments rest. He objects to the claims that 1) there can be a probability distribution over admissible values of some physical parameters when in fact those are left *indeterminate* by existing theories, and more strenuously that 2) the observed value has *low probability* instead of a *neutral* probability. Norton has suggested that the inability to distinguish between neutral and disconfirming evidence is the sign a fundamental flaw of Bayesianism, originating from the requirement of additivity. Therefore, for him, only a radically different inductive logic can adequately represent ignorance.

Norton's challenge is valid against the Bayesian who knows too much (i.e., who represents ignorance or indifference with a single uniform probability distribution). But there

[3]The median value of the distribution obtained by such anthropic prediction is about 20 times the observed value ρ_V^{obs}, whereas predictions based on existing theories are 120 orders of magnitude higher than the observed value (Pogosian et al., 2004).

are other Bayesians.

2.2 A Bayesian Failure?

2.2.1 Neutral vs. Disconfirming

Before I discuss in more detail what a representation of neutral degree of belief—i.e., a representation of indifference or ignorance—could look like, let me rephrase the problem under consideration in Bayesian terms.

The Bayesian approach to epistemology can be characterized as follows (see, e.g., Joyce, 2010, 281-282):

- Belief is not all or nothing. One can assign degrees of belief to propositions.
- These degrees obey the laws of probabilities.[4]
- Learning implies updating an initial degree of belief (called prior) to obtain a posterior. Updating a prior for hypothesis H after acquiring evidence E will involve taking conditional probabilities and applying Bayes's theorem:

  $$p(H|E) = \frac{p(H) \cdot p(E|H)}{p(E)},$$

  where $p(E|H)$ denotes the probability of E conditional on H (i.e., the probability of E given H).
- Rational agents use their graded beliefs to choose actions with higher expected value.

A further principle that is sometimes invoked as a constraint on probability assignments, the principle of indifference claims that one must assign the same probability value

[4] (1) For a probability function p, $\forall \alpha, p(\alpha) \geq 0$; (2) if α is logically true, then $p(\alpha) = 1$; (3) additivity: if α, β are incompatible ($p(\alpha \& \beta) = 0$), then $p(\alpha \vee \beta) = p(\alpha) + p(\beta)$. It follows from these laws that $\forall \alpha, p(\alpha) + p(\neg \alpha) = 1$.

to equipossible events, or events about which we are equally ignorant or indifferent. If there are many such events, additivity dictates that each of them be assigned a uniform, low probability value, which is equivalent to saying that each event is improbable (or, equivalently, that their negation is probable).

We can now see how to rephrase the problematic arguments considered in § 2.1:

1. The value of a parameter k is left indeterminate by our background knowledge. According to the principle of indifference, that indeterminateness is represented by a constant probability distribution widely spread over the admissible values of k, each of which having low probability $p(k|B)$.

2. A theory T makes the observed value k_{obs} much more probable:

$$p(k_{\text{obs}}|T\&B) >> p(k_{\text{obs}}|B).$$

3. According to Bayes's theorem, we then have

$$\frac{p(T|k_{\text{obs}}\&B)}{p(T|B)} = \frac{p(k_{\text{obs}}|T\&B)}{p(k_{\text{obs}}|B)} >> 1.$$

   In other words, observing k_{obs} lends strong support to T.

For Norton, this confirmation is unwarranted. It is based on a flaw of Bayesianism itself—namely, it has no ability to represent neutrality due to additivity (Norton, 2010, 501-502). Assigning a definite low probability value to a proposition about which we are ignorant is then *turning ignorance into improbability*. It comes down to conflating disconfirming evidence ($p(H|E) << 1$) with neutral evidential support ($p(H|E) = p(\neg H|E)$).

2.2.2 A Non-Bayesian Notion of Neutral Support?

Norton (2007a, 2008, 2010) introduced the following criteria for a candidate for representation of neutral evidential support (or indifference, or ignorance):

- it cannot be additive (and therefore does not obey the laws of probability),
- it cannot be represented by the degrees of a one-dimensional continuum, such as the reals in $[0, 1]$,
- it must be invariant under redescription,[5]
- it must be invariant under negation: if we are ignorant or indifferent as to whether or not α, we must be equally ignorant as to whether or not $\neg\alpha$.

It is clear that usual Bayesianism cannot meet all these criteria. We should then look for another framework for an inductive logic that would allow one not only to express ignorance and indifference, but also to compare credences and carry out inferences and confirmation. Norton confesses that he "know[s] of no adequate theoretical representation" of such a framework (Norton, 2010, 504). He then simply refers to this representation of neutral support as 'I', for 'indifference' or 'ignorance', with the following properties expressed in terms of a (non-probabilistic) credal function p:

- $\forall\alpha, p(\alpha) = I \rightarrow p(\neg\alpha) = I$ (invariance under negation),
- $\forall\alpha_1, \alpha_2$ mutually exclusive (but $\alpha_1 \vee \alpha_2 \not\equiv \top$), $p(\alpha_1) = p(\alpha_2) = I \rightarrow p(\alpha_1 \vee \alpha_2) = I$ (non-additivity).[6]

This framework must also preserve the values $p(\top) = 1 = 1 - p(\bot)$.[7]

This set of criteria for a representation of neutral evidential support is compelling. Because of this conflict with additivity, representing neutrality is a serious challenge for

[5] The invariance under redescription only requires that the probability value that corresponds to neutral support for a same event must not depend on how this event is described. For instance, in the example given above in § 2.1, book length was given in terms of number of words and could be redescribed in terms of number of pages or lines.

[6] Strictly speaking, it is not entirely appropriate to define this condition in terms of additivity. For a representation of credence to be 'non-additive' in the sense of interest to Norton here, it has to fulfill the following condition: $\forall\alpha, \beta$ incompatible propositions about which we are completely indifferent or ignorant, we can have neither $p(\alpha \vee \beta) > p(\alpha)$ nor $p(\alpha \vee \beta) > p(\beta)$.

[7] $\top$ is an unconditionally true statement, and $\bot$ an unconditionally false one.

Bayesianism. But I will show that we need not abandon Bayesianism altogether, and that enriched versions of it already in use can satisfy these criteria to a certain extent.

2.3 A Bayesian Notion of Neutral Support

2.3.1 Bayesian Credences Need Not Have Sharp Values

It has been argued (see, e.g., Levi, 1974; Walley, 1991; Joyce, 2010) that Bayesian credences need not have sharp values, and that there can be imprecise probabilities.[8] The difficulty to assign sharp values to credences was already raised by Kyburg (1978), who saw this as psychologically unrealistic. An imprecise probabilities model recognizes "that our beliefs should not be any more definitive or unambiguous than the evidence we have for them." (Joyce, 2010, 320)

It is possible to reject a precise Bayesian model in favor of a less exact one, in which credences are not well-defined but allow for *imprecise values*. Joyce defended an imprecise model in which credences are not represented merely by a range of values, but rather by a *family* of (probabilistic) credence functions. In this imprecise probability model,

1. a believer's overall credal state can be represented by a family C of credence functions $[c_i]$ $(\dots)$. Facts about the person's opinions correspond to properties common to all the credence functions in her credal state.

2. If the believer is rational, then every credence function in C is a probability.

3. If a person in credal state C learns that some event D obtains $(\dots)$, then her post-learning state will be $C_D = \left\{ c(X|D) = c(X)\dfrac{c(D|X)}{c(D)} : c \in C \right\}$.

[8]'Imprecise credence' is more appropriate than 'imprecise probability' since it does not necessarily obey the laws of probability. Here I nevertheless use both expressions interchangeably, as is done in the literature.

4. A rational decision-maker with credal state C is obliged to prefer one action α to another α^* when α's expected utility exceeds that of α^* relative to every credence function in C. (Joyce, 2010, 288)

In other words, in this imprecise probability model, a credal state can be represented by a family of functions that behave as usual Bayesian, probabilistic credence functions. Joyce also offers an analogy that illustrate this model: the overall credal state C acts as a committee whose members (each being analogous to a credence function c_i) are rational agents who do not all agree with each other and who all update their credence in the same way, by conditionalizing on evidence they all agree upon. With this analogy, the properties of the overall credal state C correspond to those common to all the committee members.

In order to make comparative confidence claims, different criteria can be adopted (corresponding to different decision-making rules). Depending on the criterion chosen, one will be more confident in an event than in another event if

- it has maximum lower expected value ($\Gamma-$minimax criterion),
- it has maximum higher expected value ($\Gamma-$maximax),
- it has maximum expected value for all distributions in the credal set (maximality),
- it has a higher expected value for at least one distribution in the credal set ($E-$admissibility), or
- its lower expected value on all distributions in the credal set is greater than the other event's highest expected value on all distributions (interval dominance).[9]

This model allows one to simultaneously represent sharp and imprecise credences, but also comparative probabilities. It can accommodate sharp credences, for which the

[9]This list is not exhaustive. It is beyond the scope of this paper to compare and assess these criteria. See (Troffaes, 2007; Huntley et al., 2014) for reviews.

usual condition of additivity holds. But it can also accommodate less sharply defined relationships when credences are ambiguous. It does so by means of a family of credence functions, each of which is a Bayesian function that obeys the laws of probability.

2.3.2 Neutral Support with Imprecise Credences

This imprecise model allows one to distinguish two notions of neutral support: distinguish stochastic independence (true independence of evidence for a given hypothesis) and unknown interaction (unknown dependence).

With a single (precise) probability distribution, two variables α_m and α_n in appropriate algebras are stochastically independent if $p(\alpha_m|\alpha_n) = p(\alpha_m)$ whenever $p(\alpha_n) > 0$. Different concepts have been suggested to extend the notion of stochastic independence of two variables α_m and α_n to the imprecise model (see Cozman, 2012), among which:

- complete independence, if stochastic independence of α_m and α_n obtains for each distribution $c_i \in C$,[10]

- confirmational irrelevance, if $C(\alpha_m|\alpha_n) = C(\alpha_m)$,

- epistemic irrelevance, if stochastic independence obtains for the lowest expectation among all functions $c_i \in C$, or the related, symmetric concept of epistemic independence (if α_m is epistemically irrelevant to α_n and α_n to α_m).

In contrast, unknown interaction between two variables α_m and α_n can be represented by a credal set that contains credence functions that differ on the correlation between α_m and α_n. Such a credal set would include credence functions c_i that are such that $c_i(\alpha_m \& \alpha_n) \leq c_i(\alpha_m) \cdot c_i(\alpha_n)$ and others functions c_j such that $c_j(\alpha_m \& \alpha_n) > c_j(\alpha_m) \cdot c_j(\alpha_n)$. Following the committee analogy once again, in a state of unknown interaction between α_m and α_n, jury members disagree with each other about that interaction. The

[10] As discussed in (Cozman, 2012), this definition violates convexity.

overall credal state of the committee neither favors nor dismisses any of the opinions of its jury members; it can only express a lack of agreement.

The case of complete ignorance or indifference is of particular interest for the anthropic argument presented in § 2.1: we want to represent *complete ignorance or indifference* as to what value a certain physical parameter should have. With the imprecise model, recalling the committee analogy, complete ignorance as to whether or not α (i.e., complete indifference between which of α or $\neg\alpha$ is more likely) can be represented by a committee in which there is *no agreement* among its members about whether or not α is more or less probable than $\neg\alpha$. This notion of neutral support is analog to Norton's requirement that, in order to express neutral support for an event α, a measure of belief about α be equal to a corresponding measure of *disbelief* about α (Norton, 2007a). In the context of credal sets, this requirement is fulfilled, for example, by a committee composed as follows:

> for every jury member and for every contingent event α, there exists a jury member who has as much disbelief in α as another has belief in it.

Or, to put it in terms of credence functions,

> if C is a credal state representing a family of credence functions c about contingent events α, if $\forall c, \alpha$, $\exists\, c' \in C$ such that $c'(\alpha) = 1 - c(\alpha)$, then C corresponds to a state of ignorance or indifference.[11]

If all the jury members are rational (i.e., if all the credence functions are probability functions), we must have, by definition, $C(\top) = 1 = 1 - C(\bot)$.[12]

Such a credal state about a given event α can then be updated by conditionalizing upon new evidence, taken as such by all the jury members (i.e., all the credence functions

[11] This requirement corresponds in fact more to a state of indifference than to one of ignorance. Indeed, one may argue that a credal set that gives the set of values $\{0.1, 0.8\}$ is a better representation of ignorance—but not one of indifference—than one that gives the set of values $\{0.49, 0.51\}$. I am here overlooking distinctions between these two notions.

[12] i.e., $\forall c \in C, c(\top) = 1 = 1 - c(\bot)$.

$c \in C$). Thereby, after new evidence E is gathered, the range of values taken by all the credence functions of C, $c(\alpha|E)$, is susceptible to change, and so is the overall credal state regarding α.

A trivial but extreme example of representation of *complete* ignorance by means of a set of credal functions is the set $\mathfrak{I}$ of *all* possible probability distributions. For any proposition α about which we are completely ignorant, that representation of ignorance would give us $C(\alpha) = [0, 1]$. In case of complete ignorance, excluding possible probability distributions compatible with our evidence is "pretending to have information [we do] not possess." (Joyce, 2005, 170) This proposal meets all of Norton's criteria for a representation of ignorance (see above § 2.2.2).

There is however a good reason not to be content with such an extreme representation of ignorance. Indeed, in that set $\mathfrak{I}$ of all possible probability distributions will be extremely sharp probability distributions that require an unreasonably large—or even infinite—number of updatings before they can yield posteriors distributions that are significantly different (see, e.g., Rinard, 2013, for a recent discussion). Such distributions in $\mathfrak{I}$ are said to be dogmatic, and consequently the whole set $\mathfrak{I}$ is dogmatic. A representation of complete ignorance $\mathfrak{I}$, and generally any vacuous prior, entails a vacuous posterior. This should prevent such a set from being used in an inferential process in which we may hope to move *away* from a state of ignorance after a certain number of iterations of Bayesian updating. This representation of ignorance by means of a family of credal functions, although it satisfies Norton's criteria for ignorance, is *incompatible with learning*. That is why imprecise statisticians, who are interested in inferential processes, prefer to deal with 'near-ignorance' (i.e., broad credal intervals smaller than [0,1]) rather than complete ignorance, thereby ruling out dogmatic priors (see Moral, 2012; Augustin et al., 2014; Walley, 1991, §7.3.7).

By way of example, let us represent our near-ignorance about three mutually exclusive propositions $\alpha_1, \alpha_2, \alpha_3$ with a credal set $C = \{c_i\}$ consisting of probability functions defined as follows:

	α_1	α_2	α_3
$c_1(\alpha_i)$	0.9	0.05	0.05
$c_2(\alpha_i)$	0.05	0.9	0.05
$c_3(\alpha_i)$	0.05	0.05	0.9

and for all $i, j \in \{1, 2, 3\}$, for all $x \neq \alpha_i$ or their disjunctions, $c_j(x) = 0$.

With this credal set, we cannot be more confident in any value of α_m than in any other; and for all α_m, α_n, $m \neq n$, $C(\alpha_m) = C(\alpha_n)$. A constant, unique probability distribution could express this as well. But this credal set tells us more than that, namely that, for all m, we cannot be more confident in any α_m than in its negation.

This example does not satisfy the characterization of ignorance proposed earlier. However, we can see that, with a few amendments, this set suffices to represent indifference about any of these propositions α_i, and can meet the requirements for a representation of neutral support (see above §2.2.2):

- it is not a sharp value in $[0, 1]$,
- it can be defined so as to be invariant under redescription: $\forall \alpha, a$ (a, redescription of α), if $C(\alpha)$ represents a state of ignorance, so will $C(a)$,[13]
- it is invariant under negation (we do not have more confidence in α than in its negation).
- we do have $C(\alpha_1 \vee \alpha_2 \vee \alpha_3) = C(\top) = 1 = 1 - C(\bot)$.

The criterion of non-additivity (see above note 6) cannot be satisfied in a trivial manner. In the example above, we have $\forall i, m, n \in \{1, 2, 3\}, m \neq n, c_i(\alpha_m \vee \alpha_n) > c_i(\alpha_m)$. That is also true of the bounds of the credal intervals (they will not be non-additive). But we

[13] If the functions in this set are described as Dirichlet distributions, then this criterion will be satisfied (see, e.g., de Cooman et al., 2009).

can prevent that by adding to our set credal functions c_k such that $c_k(\alpha_1 \vee \alpha_2) = c_k(\alpha_1)$ and so on. That is, we can add to our initial set three other functions c_4, c_5, and c_6 so as to have $\forall i \in \{1, 2, 3\}, \exists\, c_j \in C, j \in \{4, 5, 6\}$ such that $c_j(\alpha_i) = 0$. But the newly added functions are not reasonable, since if we agreed that the propositions α_i are contingent, then no function ruling them out completely should be accepted in our credal set.

However, adopting certain conventions can mitigate the effects of additivity. If among the criteria for comparative confidence claims with credal sets introduced earlier in § 2.3.1 we choose that of interval dominance, then our representation of credence is non-additive in this example (i.e., for all $m, n \in \{1, 2, 3\}$, we are not more confident in $\alpha_m \vee \alpha_n$ than in α_m). However, one might argue that interval dominance is often not a desirable criterion. It is a demanding criterion that allows one to express a difference in confidence between two propositions only if one is unambiguously more certain than the other. This criterion is arguably not fined-grained enough to help us for most of the inferences we are likely to encounter. Other, often more desirable comparison criteria, however, will result in additive imprecise credences. With other comparison rules, we could under circumstances circumvent additivity by adopting threshold values, beyond (respectively below) which all values are considered to be equally confirmatory (respectively disconfirmatory). For instance, in the set given above, if we consider that any value below 0.1 is equally disconfirmatory and any value beyond 0.9 is equally confirmatory, then we lose all additivity.

We could also adopt a different strategy that does not involve such conventions. For any proposition α about which we are ignorant (whether the α_i or their Boolean combinations), we can define a credal set C such that $C(\alpha) = C(\neg\alpha)$. The imprecise model allows us to treat equipossible propositions in the same way, which does not mean that they must receive a same probability value or be represented by the same credal set. By 'sameness of treatment' of contingent, equipossible events about which we are ignorant or indifferent, I mean that our credal state of ignorance about them would be modeled in the same manner, by a credal set having the same desired properties. We will come back to this in the next section.

With the imprecise model, not one single set of functions or one set of rules to define it will be suited to truly represent ignorance in all situations, unless we are ready to represent a state of complete ignorance by the undesirable and unreasonable set $\mathfrak{I}$ of all possible credal functions. But this model allows one to define sets that do not favor any of the propositions about which we are ignorant and that is suited to a particular ensemble of propositions under consideration. It does so by means of Bayesian credence functions, which allows for our credal state to be updated and evolve. All that this requires is that we do not demand that agent's credences have sharp values.

2.3.3 Norton's Objections

Interpretative objections

Norton (2007a,b, 2008) has formulated several objections to the imprecise model. He has expressed a general discomfort with what he considers to be an inadequate approach, "an attempt to simulate an inherently nonadditive logic with an additive measure, rather than to seek the logic directly." (Norton, 2010, 504, note 4). But as indirect or contrived as this method might seem, it is successfully applied in statistical analysis (see, e.g., Walley, 1991; Augustin et al., 2014). Yet Norton has raised more pointed criticisms of the imprecise model about the very question of representing ignorance:

> the representation [of ignorance by sets of probability functions] is not literally correct. That is, ignorance is not the maintaining of all possible beliefs at once; it is the maintaining of none of them. So we should regard the device of convex sets as a way of simulating ignorance through a convenient fiction.[14] (Norton, 2007b, § 4.2)

[14] This remark also applies to non-convex sets.

Or elsewhere:

> The sort of ignorance I seek to characterize is first order ignorance; it is just not knowing which is the true outcome; not a second order uncertainty about an uncertainty. (Norton, 2007a, § 6.2)

Such criticism is in fact not specific to the question of representing ignorance. It is aimed more generally at the use of *several* credal functions in order to represent a *unique* credal state.[15] But it does not apply to the imprecise model we have considered here, and, in general, proponents of imprecise probabilities need not endorse this view. Indeed, this model does not allow "the maintaining of all possible beliefs at once." Even though a credal set may be comprised of several credal functions, the agent's credence it represents is unique; its properties are those that are common to all the credal functions in that set. If all possible beliefs could be held at once—or rather, if no particular belief can be more certain than any other—an agent's credence would not be multiple, it would just *not be any* of these particular beliefs.

Indifference and self-duality

Another criticism more specifically aimed at the issue of representing ignorance or indifference deserves a closer examination. Norton (2007a) argues that a representation of ignorance should satisfy what he calls a condition of self-duality. The dual of a measure of belief is a measure of disbelief. If we are ignorant about a proposition α, our degree of belief that α should not be different from our degree of disbelief that α:

> An epistemic state of complete ignorance is invariant in its contingent propositions (. . .); that is, the state is self-dual in its contingent propositions, so

[15] The following passage makes it clear that Norton thinks of the use of a set of probability functions as allowing the simultaneous representation of several states of belief: "the use of sets renders ignorance as a second order sort of belief. We allow that many different belief-disbelief states are possible. We represent ignorance by presenting them all, in effect saying that we dont know which is the pertinent one." (Norton, 2007a, § 6.2, 248)

> that $m(\alpha) = M(\alpha) = m(\neg\alpha)$ for all contingent α [where $m(\alpha)$ is a measure of the belief that α and $M(\alpha)$ its dual]. (Norton, 2007a, § 6.1, 247)

This follows from the requirement of invariance under negation for a representation of neutral support.

We saw earlier in § 2.3.2 that no single probability function (or in general no measure, additive by definition) can meet this criterion.[16] Any probability function will necessarily express a (non-strictly) increasing belief as we go from propositions to their logical consequences (e.g., for any functions c in a credal set, $\alpha_1 \vdash \alpha_1 \vee \alpha_2 \rightarrow c(\alpha_1 \vee \alpha_2) \geq c(\alpha_1)$). On the other hand, a measure of disbelief should be (non-strictly) *decreasing* as we go from propositions to their logical consequences. Consequently, no probability function can simultaneously express our belief and our disbelief in a certain proposition, even if we consider only contingent propositions. This, however, is not necessarily true of credal sets. We also saw in § 2.3.2 that if comparative confidence claims are based on interval dominance, then imprecise credences need not be additive for contingent propositions. Consequently, an imprecise credence represented by a set of functions as the one given in § 2.3.2 can simultaneously represent belief and disbelief (at least for contingent propositions).

However, on Norton's account (Norton, 2007a, §§ 3.2, 6.2) a credal set C representing ignorance is self-dual if it contains probability functions c_i and their duals c_i' such that $\forall c_i \in C, c_i'$ is such that $\forall \alpha, c_i(\alpha) = c_i'(\neg\alpha)$ and $c_i(\neg\alpha) = c_i'(\alpha)$. But, as we just saw, the duals of probability functions cannot be probability functions, and consequently a set containing probability functions and their duals cannot comprise probability functions only.

In contrast, for any given set of contingent propositions, the imprecise model allows us to define sets that represent our indifference or ignorance. We saw above in § 2.3.2 a proposal for a representation of neutral support, according to which if C is a credal

[16] Unless we are dealing with only two mutually exclusive propositions.

state representing a family of credence functions c about contingent events α, C corresponds to a state of indifference if $\forall c, \alpha$, $\exists\, c' \in C$ such that $c'(\alpha) = 1 - c(\alpha)$. This proposal differs from Norton's criterion of self-duality (note the different placement of the universal quantifier on α). The imprecise model as I have considered it here does not guarantee us than once a credal set is defined so as to represent indifference about a set of propositions, it will necessarily adequately represent our indifference about all their Boolean combinations, nor, in general, that it will adequately provide neutral support to any other proposition we are indifferent about. This model does not suggest either that the same rules used to define a credal set representing indifference should apply to all situations (this will be illustrated below in § 2.4). But for any given set of contingent propositions—for instance, some propositions and their negations—we can construct a credal set that represents our indifference; thus, it will represent our indifference about these propositions and the negation of each of them. For these propositions, it is reasonable to expect this sort of duality from our representation of belief (i.e, from the credal set as a whole), but not from its components.

If we consider only contingent propositions, and if comparison in our level of confidence is based on interval dominance, then with imprecise probabilities we have at our disposal a representation of ignorance or indifference that shares with a self-dual representation the relevant properties for a representation of neutral support, namely the ability to simultaneously represent belief and disbelief. With other comparison criteria (e.g., Γ-minimax), a self-dual representation of ignorance or indifference as one Norton demands could only express nothing less than *complete* ignorance. We saw in § 2.3.2 that such a credal set, $\mathfrak{I}$, exists, but we also saw that it is incompatible with learning. The less demanding representations of ignorance or indifference I have considered here have over a self-dual representation a clear expressive advantage. We saw that the imprecise model allows one to distinguish between stochastic independence, epistemic irrelevance, unknown interaction, or ignorance or indifference about the value of a parameter. It can fulfill criteria of a representation of ignorance better than what a single probability function can do, yet it does so by means of probability functions, each of which can subsequently be modified following Bayes's rule.[17] Consider for instance the 3-function example from § 2.3.2. Now assume that the probability of α_i in that example corresponds to the probability of drawing the numbered ball B_i from an urn (with replacement between each draw). Assume further that we are interested in determining a bias—of our urn or the balls—in this drawing. How should we represent our initial state of ignorance in a way that does not prevent us from eventually finding a value after a sufficient iteration of Bayesian updating? It is unclear how Norton's non-probabilistic, self-dual representation of complete ignorance, 'I', can evolve, whereas it is possible for our 3-function credal set to yield such a result eventually.[18] Notwithstanding the lack of self-duality, the ability for credal sets representing indifference or ignorance to be updated is a desirable feature that makes an imprecise representation of indifference a more interesting element of inductive logic than Norton's self-dual measure I.

The kind of representation of ignorance that Norton seeks is part of a larger inductive logic yet to be carried out. It is in the light of the search for a unique representation of our credences that Norton imposes this criterion of self-duality of complete ignorance, applicable to *any* proposition about which we are ignorant. In this context, it is a plausible requirement. But it is a very strong condition that can leave one wondering how a unique representation as the one Norton proposes—which would yield the same value 'I' whatever the event about which we are ignorant—can be used in a fruitful inductive process. Norton offers no compelling reason why his demanding notion of self-duality should be required of alternative representations of ignorance or indifference such as the ones the imprecise model affords us.

[17]Further discussion about the expressive advantages of imprecise probabilities can be found in (de Cooman and Miranda, 2007; Miranda and de Cooman, 2014).

[18]See (Piatti et al., 2009; Moral, 2012) for a discussion about the conditions on a near-ignorance credal set required for learning. These requirements favor the use of Dirichlet distributions.

2.4 Dissolving Cosmic Confusions

In order to emphasize how ill-equipped Bayesianism (or any other inductive framework based on additive measures) is in order to deal with ignorance, and to show its inability to avoid unwarranted conclusions, Norton (2010) invited us to consider two instances of 'cosmic confusion.' These are anthropic reasoning (see above § 2.1) and the Doomsday argument. In both cases, we draw conclusions from a *lack of knowledge*. We can see how the use of imprecise credences can dissolve what Norton referred to as 'cosmic confusions' in a way no unique probability distribution can.

The Doomsday argument is a family of arguments about humanity's likely survival (see, e.g., (Bostrom, 2002, §§ 6–7), (Richmond, 2006) for reviews). It allows one to compare the likelihood of two scenarios about humanity's survival or even make a prediction about the end date for humanity based only on the assumption that our place on humanity's timeline is random. Some versions of this argument have been addressed within the framework of orthodox Bayesianism so as to block any conclusion we could draw from these assumptions alone. But a variant of this argument, based on the assumption that we have a random birth rank among all humans (Gott, 1994), cannot be dissolved without appealing to imprecise credences. The argument goes as follows. Let r be my birth rank (i.e., I am the r^{th} human to be born), and N the total number of humans that will ever be born.

1. Assume that there is nothing special about my rank r. Following the principle of indifference, whatever r, the probability of r conditional on N is $p(r|N) = \dfrac{1}{N}$.
2. Assume the following improper prior[19] for N: $p(N) = \dfrac{k}{N}$. k is a normalizing constant whose value does not matter.

[19]As Gott (1994) recalls, this choice of prior is fairly standard (albeit contentious) in statistical analysis. It is the Jeffreys prior for the unbounded parameter N, such that $p(N)\,\mathrm{d}N \propto \mathrm{d}\ln N \propto \dfrac{\mathrm{d}N}{N}$. This means that the probability for N to be in any logarithmic interval is the same. This prior is called improper because it is not normalizable, and it is usually argued that it is justified when it yields a normalizable posterior.

3. This choice of distributions $p(r|N)$ and $p(N)$ gives us the prior distribution $p(r)$, with $N \geq r$:[20]

$$p(r) = \int_{N=r}^{N=\infty} p(r|N)p(N)\,\mathrm{d}N = \int_{N=r}^{N=\infty} \frac{k}{N^2}\,\mathrm{d}N = \frac{k}{r}.$$

4. Then, Bayes's theorem gives us

$$p(N|r) = \frac{p(r|N)\cdot p(N)}{p(r)} = \frac{r}{N^2},$$

which favors small N and allows us to make an estimate for N at any confidence-level.

This result should strike us as surprising: we should not be able to learn something from nothing! Indeed, according to that argument, we can make a prediction for N based only on knowing our rank r and on *not* knowing anything about the probability of r conditional on N, i.e., on being indifferent—or equally uncommitted—about any value it may take.

In this argument, any choice of prior probability distribution will result in a prediction for N, at any confidence-level. However, if our prior ignorance or indifference about N, $C(N)$, is represented by a credal set containing an *infinity* of credal functions, $\{c : c \in C\}$, each normalizable, defined on $\mathbb{N}^{>0}$, and such that $\forall c \in C, \lim_{N\to\infty}(c(N)) = 0$ (e.g., a family of Pareto distributions), then the resulting prediction for N diverges. In other words, this imprecise representation of prior credence in N, reflecting our ignorance about N, does not yield any prediction about N. Without the possibility for my prior credence to be represented not by a single probability distribution, but instead by an *infinite set* of probability distributions, I cannot avoid obtaining an arbitrarily precise prediction.

In the case of the cosmological constant problem (see above § 2.1), representing our

[20] I use a continuous distribution as an approximation for the discrete case.

prior ignorance or indifference about the value of the vacuum energy density ρ_V by an imprecise credal set can limit, if not entirely dissolve, the appeal of anthropic considerations. As we saw earlier, Weinberg (1987) argued that, in the absence of useful theoretical background, it was reasonable to assume a constant, uniform prior probability distribution for ρ_V within the anthropically allowed range, and then conditionalize on the number of observers each value of ρ_V would allow for. With the imprecise model, a state of indifference between different values of ρ_V within the anthropically allowed range can be expressed by a set of probability distributions $\{c_\star : c_\star \in C_\star\}$, all of which normalizable over the anthropic range and such that $\forall \rho_V, \exists\, c_{\star i}, c_{\star j} \in C_\star$ such that ρ_V is favored by $c_{\star i}$ and not by $c_{\star j}$.[21] It is in principle possible to define this prior credal set so that for any value of ρ_V, the lowest expectation value among the the posteriors is lower than the highest expectation value among the priors. If then we adopt interval dominance as a criterion for comparative confidence claims, then no observation of ρ_V will be able to lend support to our anthropic prediction.

As we saw earlier, one may object to the adoption of interval dominance in such a case. This choice of demanding confidence comparison rule could be motivated by the fact that we have no plausible alternative theoretical framework to the anthropic argument. In this context, it can be reasonable to agree to increase one's credence about the anthropic explanation *only if* it does better than any other yet unknown alternative might have done. Nonetheless, if we adopt other confidence comparison rules, it is possible with the imprecise model to construct prior credal sets that define a large interval over the anthropic range such that the confirmatory boost obtained after observing ρ_V is not nearly as vindicative as it is with a single, uniform distribution.

[21] For reasons expressed earlier in footnotes 13 and 18, this should preferably be done by means of Dirichlet distributions.

2.5 Conclusion

Norton (2010) has correctly argued that representing neutrality with a broadly spread single probability distribution amounts to conflating ignorance with improbability. He has shown how this leads to unwarranted confirmations. I here claim that a credal state of ignorance should best be represented by an imprecise credence. With this approach, merely acquiring information about the value of a certain parameter cannot suffice to justify a sense of surprise. The imprecise model offers us a more adequate representation of neutrality and prevents prior credences from doing too much inductive work, as is illustrated by its ability to block the consequence of the Doomsday argument better than what orthodox Bayesianism can do.

We saw that, if we adopt interval dominance as a criterion for confidence comparison, it is possible for an imprecise representation of indifference to meet the criteria for a representation of neutral support put forth in (Norton, 2007a, 2008, 2010). But if we adopt less demanding confidence comparison rules, we can still construct a representation of indifference by means of credal sets that meet these criteria to a large extent, and we can do so without compromising Bayesianism altogether. It only requires that we do not demand that credences be sharp nor that a unique representation be applicable to all cases of ignorance or indifference (i.e., that self-duality be abandoned).

One can see Norton's argument as emphasizing the perils of excessive structure imported from probability theory into inductive logic, and then arguing that we need to eliminate much of that structure. There are several ways to modify the mathematical structure to counter the assumption of additivity. Norton's stated motivations are not sufficient to force us to adopt the framework he advocates. The imprecise model has the advantage of allowing us to distinguish different kinds of ignorance and indifference. More importantly, it makes it possible to move out of a state of ignorance when we acquire interesting information.

Bibliography

Aguirre, A. (2001). Cold Big-Bang Cosmology as a Counterexample to Several Anthropic Arguments. *Physical Review D*, 64(8):1–12.

Augustin, T., Coolen, F. P., de Cooman, G., and Troffaes, M. C., editors (2014). *Introduction to Imprecise Probabilities*. Wiley & Sons.

Bostrom, N. (2002). *Anthropic Bias: Observation Selection Effects in Science and Philosophy*. Routledge.

Cozman, F. G. (2012). Sets of Probability Distributions, Independence, and Convexity. *Synthese*, 186(2):577–600.

de Cooman, G. and Miranda, E. (2007). Symmetry of Models versus Models of Symmetry. In Harper, W. L. and Wheeler, G., editors, *Probability and Inference. Essays in Honour of Henry E. Kyburg Jr*, pages 67–149. College Publications, London.

de Cooman, G., Miranda, E., and Quaeghebeur, E. (2009). Representation Insensitivity in Immediate Prediction under Exchangeability. *International Journal of Approximate Reasoning*, 50(2):204–216.

Gott, J. R. (1994). Future Prospects Discussed. *Nature*, 368(March):108.

Huntley, N., Hable, R., and Troffaes, M. C. (2014). Decision Making. In Augustin, T., Coolen, F., and Troffaes, M. C., editors, *Introduction to Imprecise Probabilities*, pages 190–206. Wiley & Sons, Chichester, U.K.

Joyce, J. M. (2005). How Probabilities Reflect Evidence. *Philosophical Perspectives*, 19:153–178.

Joyce, J. M. (2010). A Defense of Imprecise Credences in Inference and Decision Making. *Philosophical Perspectives*, 24(1):281–323.

Kyburg, E. H. (1978). Subjective Probability: Criticisms, Reflections, and Problems. *Journal of Philosophical Logic*, 7(1):157–180.

Levi, I. (1974). On Indeterminate Probabilities. *The Journal of Philosophy*, 71(13):391–418.

Miranda, E. and de Cooman, G. (2014). Structural Judgments. In Augustin, T., Coolen, F., de Cooman, G., and Troffaes, M. C., editors, *Introduction to Imprecise Probabilities*, pages 56–78. Wiley & Sons, Chichester, U.K.

Moral, S. (2012). Imprecise Probabilities for Representing Ignorance about a Parameter. *International Journal of Approximate Reasoning*, 53(3):347–362.

Norton, J. D. (2007a). Disbelief as the Dual of Belief. *International Studies in the Philosophy of Science*, 21(3):231–252.

Norton, J. D. (2007b). Probability Disassembled. *The British Journal for the Philosophy of Science*, 58(2):141–171.

Norton, J. D. (2008). Ignorance and Indifference. *Philosophy of Science*, 75(1):45–68.

Norton, J. D. (2010). Cosmic Confusions: Not Supporting versus Supporting Not. *Philosophy of Science*, 77(4):501–523.

Piatti, A., Zaffalon, M., Trojani, F., and Hutter, M. (2009). Limits of Learning about a Categorical Latent Variable under Prior Near-Ignorance. *International Journal of Approximate Reasoning*, 50(4):597–611.

Pogosian, L., Vilenkin, A., and Tegmark, M. (2004). Anthropic Predictions for Vacuum Energy and Neutrino Masses. *Journal of Cosmology and Astroparticle Physics*, 7(005):1–17.

Richmond, A. (2006). The Doomsday Argument. *Philosophical Books*, 47(2):129–142.

Rinard, S. (2013). Against Radical Credal Imprecision. *Thought: A Journal of Philosophy*, 2(2):157–165.

Troffaes, M. C. (2007). Decision Making under Uncertainty Using Imprecise Probabilities. *International Journal of Approximate Reasoning*, 45(1):17–29.

Vilenkin, A. (1995). Predictions from Quantum Cosmology. *Physical Review Letters*, 74(6):4–7.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London.

Weinberg, S. (1987). Anthropic Bound on the Cosmological Constant. *Physical Review Letters*, 59(22):2607–2610.

Weinberg, S. (2000). A Priori Probability Distribution of the Cosmological Constant. *arXiv preprint astro-ph/0002387*, pages 0–15.

Weinberg, S. (2007). Living in the Multiverse. In Carr, B., editor, *Universe or Multiverse?*, chapter 2, pages 29–42. Cambridge University Press.

Chapter 3

Blurring Out Cosmic Puzzles[1]

3.1 Introduction

The Doomsday argument and the appeal to anthropic bounds to solve the cosmological constant problem are two examples of puzzles of probabilistic confirmation. These arguments both make ‘cosmic’ predictions: the former gives us a probable end date for humanity, and the second a probable value of the vacuum energy density of the universe. They both seem to allow one to draw unwarranted conclusions from a *lack of knowledge*, and yet one way of formulating them makes them a straightforward application of Bayesianism. They call for a framework of inductive logic that allows one to represent ignorance better than what can be achieved by a Bayesian approach that represents credal states by single credence functions so as to block these conclusions.

3.1.1 The Doomsday Argument

The Doomsday argument is a family of arguments about humanity’s likely survival.[2] There are mainly two versions of the argument discussed in the literature, both of which appeal to a form of Copernican principle (or principle of typicality or mediocrity). A first version of the argument endorsed by, e.g., John Leslie (1990) dictates a probability

[1]This chapter was presented at the 2014 Biennial Meeting of the Philosophy of Science Association, and it is forthcoming in the corresponding issue of *Philosophy of Science*.

[2]See, e.g., (Richmond, 2006; Bostrom, 2002, §§ 6–7) for reviews.

shift in favor of theories that predict earlier end dates for our species assuming that we are a typical—rather than atypical—member of that group.

The other main version of the argument, referred to as the "delta-t argument," was given by Richard Gott (1993) and has provoked both outrage and genuine scientific interest.[3] It claims to allow one to make a prediction about the total duration of any process of indefinite duration based only on the assumption that the moment of observation is randomly selected. A variant of this argument, which gives equivalent predictions, reasons in terms of random selection of one's rank in a sequential process (Gott, 1994).[4] The argument goes as follows:

Let r be my birth rank (i.e., I am the r^{th} human to be born), and N the total number of humans that will ever be born.

1. Assume that there is nothing special about my rank r. Following the principle of indifference, for all r, the probability of r conditional on N is $p(r|N) = \dfrac{1}{N}$.

2. Assume the following improper prior probability distribution for N: $p(N) = \dfrac{k}{N}$. k is a normalizing constant, whose value doesn't matter.

3. This choice of distributions $p(r|N)$ and $p(N)$ gives us the prior distribution $p(r)$:[5] $p(r) = \displaystyle\int_{N=r}^{N=\infty} p(r|N)p(N)\,\mathrm{d}N = \int_{N=r}^{N=\infty} \frac{k}{N^2}\,\mathrm{d}N = \frac{k}{r}$.

4. Then, Bayes' theorem gives us $p(N|r) = \dfrac{p(r|N) \cdot p(N)}{p(r)} = \dfrac{r}{N^2}$, which favors small N.

The choice of the Jeffreys prior for the unbounded parameter N in step 2 is such that the probability for N to be in any logarithmic interval is the same; that is, we have

[3] See, e.g., (Goodman, 1994) for opprobrium and (Wells, 2009; Griffiths and Tenenbaum, 2006) for praise.

[4] The latter version doesn't violate the reflection principle—entailed by conditionalization—according to which an agent ought to have now a certain credence in a given proposition if she is certain she will have it at a later time (Monton and Roush, 2001).

[5] I use a continuous distribution as an approximation for the discrete case.

$p(N)\,\mathrm{d}N \propto \mathrm{d}\ln N \propto \dfrac{\mathrm{d}N}{N}$. This prior is called improper because it is not normalizable, and it is sometimes argued that it is justified when it yields a normalizable posterior. Although this is a contentious assumption, we will see that no other precise distribution would allow us to avoid the conclusion of the Doomsday argument.

To find an estimate with a confidence α, we solve $p(N \leq x|r) = \alpha$ for x, with $p(N \leq x|r) = \int_r^x p(N|r)\,\mathrm{d}N$. Upon learning r, we are able to make a prediction about N with a 95%-level confidence. Here, we have $p(N \leq 20r|r) = 0.95$. That is, we have $p(N > 20r|r) < 5\%$.

According to that argument, we can make a prediction for N based only on knowing our rank r and on being indifferent about any value r conditional on N may take. We should be troubled by the fact that we can get so much information out of so little. If N is unbounded, an appeal to our typical position shouldn't allow us to make any prediction at all, and yet it does.

3.1.2 Anthropic Reasoning in Cosmology

Another probabilistic argument that claims to allow one to make a prediction from a lack of knowledge is commonly used in cosmology, in particular to solve the cosmological constant problem (i.e., explain the value of the vacuum energy density ρ_V). This parameter presents physicists with two main problems:[6]

1. The time coincidence problem: we happen to live at the brief epoch—by cosmological standards—of the universe's history when it is possible to witness the transition from the domination of matter and radiation to vacuum energy ($\rho_M \sim \rho_V$).

2. There is a large discrepancy—of 120 order of magnitudes—between the (very small) observed values of ρ_V and the (very large) values suggested by particle-physics models.

[6] See (Carroll, 2001; Solà, 2013) for an overview of the cosmological constant problem.

Anthropic selection effects (i.e., our sampling bias as observers existing at a certain time and place and in a universe that must allow life) have been used to explain both problems. Anthropic selection effects make the coincidence less unexpected, and account for the discrepancy between observations and possible expectations from available theoretical background. But there is no known reason why having $\rho_M \sim \rho_V$ should matter to the advent of life.

Steven Weinberg and his collaborators (Weinberg, 1987, 2000; Martel et al., 1998), among others, proposed that, in the absence of satisfying explanations, anthropic considerations could play a strong, predictive role. The idea is that we should conditionalize the probability of different values of ρ_V on the number of observers (or a proxy, such as the number of galaxies) taken as a function of that parameter. The probability measure for ρ_V is then $\mathrm{d}\, p(\rho_V) = \nu(\rho_V) \cdot p_\star(\rho_V)\, \mathrm{d}\rho_V$, where $p_\star(\rho)\, \mathrm{d}\rho_V$ is the prior probability distribution, and $\nu(\rho_V)$ the average number of galaxies which form for ρ_V.

By assuming that there is no known reason why the likelihood of ρ_V should be special at the observed value, and because the allowed range of ρ_V is very far from what we would expect from available theories, Weinberg and his collaborators argued that it is reasonable to assume that the prior probability distribution is constant within the anthropically allowed range, so that $\mathrm{d}p(\rho_V)$ can be calculated as proportional to $\nu(\rho_V)\, \mathrm{d}\rho_V$ (Weinberg, 2000, 2). Weinberg then predicted that the value of ρ_V would be close to the mean value in that range (assumed to yield the largest number of observers). This "principle of mediocrity," as Alexander Vilenkin (1995) called it, assumes that we are typical observers.

Thus, anthropic considerations not only help establish the prior probability distribution for ρ_V by providing bounds, but they also allow one to make a prediction regarding its observed value. This method has yielded predictions for ρ_V only a few orders of magnitudes apart from the observed value.[7] This improvement—from 120 orders of magnitude to only a few—has been seen by their proponents as vindicating anthropically-based

[7]The median value of the distribution obtained by such anthropic prediction is about 20 times the observed value ρ_V^{obs} (Pogosian et al., 2004).

approaches (see, e.g., Weinberg, 2007).

3.1.3 The Problem: *Ex Nihilo Nihil Fit*

The Doomsday argument and anthropic reasoning share a similar structure: 1) a uniform prior probability distribution reflects an initial state of ignorance or indifference, and 2) an appeal to typicality or mediocrity is used to make a prediction. This is puzzling: these two assumptions of indifference and typicality are meant to express neutrality, and yet from them alone we seem to be getting a lot of information. But assuming neutrality *alone* should not allow us to learn anything!

If anthropic considerations were only able to provide us with one bound (either lower or upper bound), then the argument used to make a prediction about the vacuum energy density ρ_V would be analogous to Gott's 1993 delta-t argument: without knowing anything about, say, a parameter's upper bound, a uniform prior probability distribution over all possible ranges and the appeal to typicality of the observed value favors lower values for that parameter.

I will briefly review several approaches taken to dispute the validity of the results obtained from these arguments. We will see that dropping the assumption of typicality isn't enough to avoid these paradoxical conclusions. I will show that, when dealing with events we are completely ignorant or indifferent about, one can use an imprecise, Bayesian-friendly framework that better handles ignorance or indifference.

3.2 Typicality, Indifference, Neutrality

3.2.1 How Crucial to Those Arguments Is the Assumption of Typicality?

The appeal to typicality is central to Gott's delta-t argument, Leslie's version of the Doomsday argument, and Weinberg's prediction. This assumption has generated much of the philosophical discussion about the Doomsday argument in particular. Nick Bostrom (2002) offered a challenge to what he calls the Self-Sampling Assumption (SSA), according to which "one should reason as if one were a random sample from the set of all observers in one's reference class." In order to avoid the consequence of the Doomsday argument, Bostrom suggested to adopt what he calls the Self-Indicating Assumption (SIA): "Given the fact that you exist, you should (other things equal) favor hypotheses according to which many observers exist over hypotheses on which few observers exist." (Bostrom, 2002) But as he noted himself (Bostrom, 2002, 122-126), this SIA is not acceptable as a general principle. Indeed, as Dennis Dieks summarized: "Such a principle would entail, e.g., the unpalatable conclusion that armchair philosophizing would suffice for deciding between cosmological models that predict vastly different chances for the development of human civilization.The infinity of the universe would become certain a priori." (Dieks, 2007, 431)

The biggest problem with Doomsday-type arguments resting on the SSA is that their conclusion depends on the choice of reference class. What constitutes "one's reference class" seems entirely arbitrary or ill-defined: is my reference class that of all humans, mammals, philosophers, etc.? Anthropic predictions can be the object of a similar criticism: the value of the cosmological constant most favorable to the advent of life (as we know it) may not be the same as that most favorable to the existence of intelligent observers, which might be definable in different ways.

Relatedly, Radford Neal (2006) argued that conditionalizing on non-indexical information (i.e., all the information at the disposal of the agent formulating the Doomsday argument, including all their memories) reproduces the effects of assuming both SSA and SIA. Conditionalizing on the probability that an observer with all their non-indexical information exists (which is higher for a later Doomsday, and highest if there is no Doomsday at all) blocks the consequence of the Doomsday argument without invoking such ad hoc principles, and avoids the reference-class problem (see also Dieks, 1992).

Although full non-indexical conditioning cancels out the effects of Leslie's Doomsday argument (and, similarly, anthropic predictions), it is not clear that it also allows one to avoid the conclusion of Gott's version of the Doomsday argument. Neal (2006, 20) dismisses Gott's argument because it rests *only* on an "unsupported" assumption of typicality. There are indeed no good reasons to endorse typicality a priori (see, e.g., Hartle and Srednicki, 2007). One might then hope that not assuming typicality would suffice to dissolve these cosmic puzzles. Irit Maor et al. (2008) showed for instance that without it, anthropic considerations don't allow one to really make predictions about the cosmological constant, beyond just providing unsurprising boundaries, namely, that the value of the cosmological constant must be such that life is possible.

My approach in this paper, however, will not be to question the assumption of typicality. Indeed, in Gott's version of the Doomsday argument given in § 3.1.1, we would obtain a prediction *even if we didn't assume typicality*. Instead of assuming a flat probability distribution for our rank r conditional on the total number of humans N, $p(r|N) = \frac{1}{N}$, let's assume a non-uniform distribution. For instance, let's assume a distribution that favors our being born in humanity's timeline's first decile (i.e., one that peaks around $r = 0.1 \times N$). We would then obtain a different prediction for N than if we had assumed one that peaks around $r = 0.9 \times N$. This reasoning, however, yields an unsatisfying result if taken to the limit: if we assume a likelihood probability distribution for r conditional on N sharply peaked at $r = 0$, we would *still* obtain a prediction for N upon learning r

(see Figure 3.1).[8]

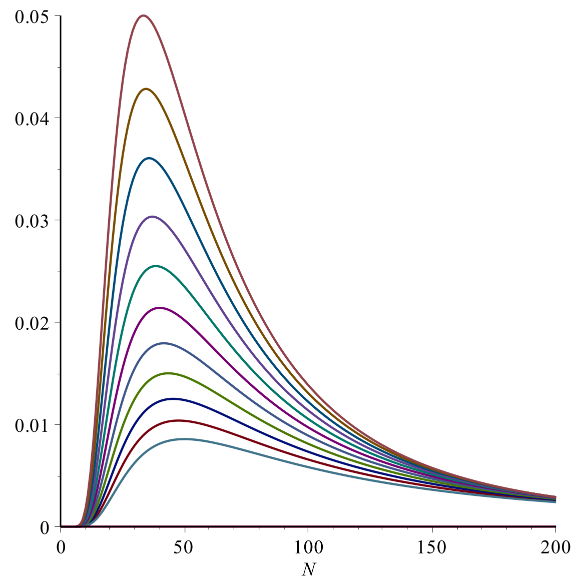

Figure 3.1: Posterior probability distributions for N conditional on r, obtained for $r = 100$ and assuming different likelihood distributions for r conditional on N (i.e., with different assumptions as to our relative place in humanity's timeline), which each peaks at different values $\tau = \frac{r}{N}$. The lowermost curve corresponds to a likelihood distribution that peaks at $\tau \to 0$, i.e., if we assume $N \to \infty$.

Therefore, in Gott's Doomsday argument, we would obtain a prediction at any confidence-level, whatever assumption we make as to our typicality or atypicality, and we would even obtain one if we assume $N \to \infty$. Consequently, it is toward the question of a probabilistic representation of ignorance or indifference that I will now turn my attention.

3.2.2 A Neutral Principle of Indifference?

One could hope that a more adequate prior probability distribution—one that better reflects our ignorance and is normalizable—may prevent the conclusion of these cosmic puzzles (especially Gott's Doomsday argument). The idea that a uniform probability distribution is not a satisfying representation of ignorance is nothing new; this discussion

[8]Tegmark and Bostrom (2005) used a similar reasoning to derive an upper bound on the date of a Doomsday catastrophe.

is as old as the principle of indifference itself.[9] As argued by John Norton (2010), a uniform probability distribution is unable to fulfill invariance requirements that one should expect of a representation of ignorance or indifference:

- non-additivity,
- invariance under redescription,
- invariance under negation: if we are ignorant or indifferent as to whether or not α, we must be equally ignorant as to whether or not $\neg\alpha$.[10]

For instance, in the case of the cosmological constant problem, if we adopt a uniform probability distribution for the value of the vacuum energy density ρ_V over an anthropically-allowed range of length μ, then we are committed to assert for instance that ρ_V is 3 times more likely to be found in a any range of length $\frac{\mu}{3}$ than in any other range of length $\frac{\mu}{9}$. This is very different from indifference or ignorance, hence the requirement of non-additivity for a representation of ignorance.

These criteria for a representation of ignorance or indifference cast doubt on the possibility for a probabilistic logic of induction to overcome these limitations.[11] I will argue that an imprecise model of Bayesianism, in which our credences can be ambiguous, will be able to explain away these problems without abandoning Bayesianism altogether.

3.3 Dissolving the Puzzles with Imprecise Credence

3.3.1 Imprecise Credence

Bayesian probability generally operates under the assumption that an agent can represent her credence by a single, sharp numerical value between 0 and 1. A common gripe

[9] See, e.g., (Syversveen, 1998) for a short review on the problem of representing non-informative priors.

[10] For an extended discussion about criteria for a representation of ignorance—with imprecise probabilities in particular—see (de Cooman and Miranda, 2007, §§ 4–5). See also Chapter 2 above.

[11] The same goes for improper priors, as was argued, e.g., by Dawid et al. (1973).

against Bayesian approaches is that this assumption is psychologically unrealistic (see, e.g., Kyburg, 1978). Moreover, for those who think of probabilities in terms of betting behavior, it would be more realistic to deal with an interval of betting prices (bounded by a selling price and a buying price), rather than a unique value (see Smith, 1961).

In a model of imprecise credences (or 'imprecise probabilities' by misuse of language) developed and defended by, e.g., Walley (1991); Joyce (2010), credences are not represented merely by a range of values, but rather by a *family* of probabilistic credence functions. In this model, an agent's credal state can be represented by a family C of probabilistic credence functions $[c_i]$, whose properties are those *common to all the credence functions* in this credal state. On this account, one's credal state upon learning that a certain event D obtains is the set of the updated credence functions $C_D = \left\{ c(X|D) = c(X)\dfrac{c(D|X)}{c(D)} : c \in C \right\}$.

In this model, each credal function (i.e., each member of a family of function that represents an agent's credal state) is treated as in a Bayesian approach that represents credal states by single credence functions. Precise probabilities are therefore a special case of the imprecise probabilities model.

Different criteria for making comparative confidence claims exist in the literature: for instance, we can say that one will be more confident in an event than in another event if

- it has maximum lower expected value ($\Gamma-$minimax criterion),
- it has maximum higher expected value ($\Gamma-$maximax),
- it has maximum expected value for all distributions in the credal set (maximality),
- it has a higher expected value for at least one distribution in the credal set ($E-$admissibility), or
- its lower expected value on all distributions in the credal set is greater than the other event's highest expected value on all distributions (interval dominance).[12]

[12] This list is not exhaustive, see (Troffaes, 2007; Huntley et al., 2014) for reviews.

This imprecise model is interesting when it comes to representing ignorance or indifference: it can do so with a *set of functions that disagree with each other.* If the agent is a committee whose members' opinions correspond to the credal functions that constitute the agent's credal state (i.e., the whole set), then this situation corresponds to one of indecision resulting from the disagreement between the committee members. How this indecision arises will depend on which of the above rules we adopt.

3.3.2 Blurring Out Gott's Doomsday Argument: Apocalypse Not Now

Let us see how we can reframe Gott's Doomsday argument with an imprecise prior credence for the total number of humans N, or more generally for the length of any process of indefinite duration X. Let our prior credence in X, C_X, be represented by a family of credal functions $\{c_\gamma : c_\gamma \in C_X\}$, each normalizable and defined on $\mathbb{R}^{>0}$. Thus, we avoid improper prior distributions. All we assume is that X is finite but can be indefinitely large. We have no reason to exclude from our prior credal set C_X any distribution that is monotonically decreasing and such that $\forall c_\gamma \in C_X, \lim_{X\to\infty}(c_\gamma(X)) = 0$.[13] Let then our prior credence consist in the following set of functions, all of which decrease but not at the same rate (i.e., similar to a family of Pareto distributions), $\left\{c_\gamma(X) = \frac{k_\gamma}{X^\gamma} : c_\gamma \in C_X\right\}$, with $\gamma > 1$ and k_γ a normalizing constant such that $k_\gamma = \frac{1}{\int_0^\infty \frac{\mathrm{d}X}{X^\gamma}}$. The limiting case $\gamma \to 1$ corresponds to $X \to \infty$, but $\gamma = 1$ must be excluded to avoid a non-normalizable distribution.

If we don't want to assume anything about the distributions in C_X (other than their being monotonically decreasing), this prior set must be such that it contains functions of decreasing rates that are arbitrarily small. That is, $\forall X \in \mathbb{R}^{>0}, \forall \epsilon \in \mathbb{R}^{<0}$, $\exists\, c_\gamma \in C_X$ such that $\frac{\mathrm{d}c_\gamma(X)}{\mathrm{d}X} > \epsilon$. This requirement applies not to any of the functions in C_X but

[13] In order to avoid too sharply peaked distributions (at $X \to 0$), constraints can be placed on the variance of the distributions (namely, an lower bound on the variance), without it affecting my argument.

to the set as a whole.

Following the steps of the argument given above in § 3.1.1, we obtain the following expression for the distributions in the credal set $\{c_\gamma(r) : c_\gamma(r) \in C_r\}$ representing our prior credence in r:

$$c_\gamma(r) = \int_{N=r}^{N=\infty} p(r|N) \cdot c_\gamma(N)\,\mathrm{d}N = \int_{N=r}^{N=\infty} \frac{k_\gamma}{N^{\gamma+1}}\,\mathrm{d}N.$$

Bayes' theorem then yields an expression for the posterior credal functions in $C_{N|r}$:

$$c_\gamma(N|r) = \frac{p(r|N) \cdot c_\gamma(N)}{c_\gamma(r)} = \frac{k_\gamma}{N^{\gamma+1} \cdot \int_{N=r}^{N=\infty} \frac{k_\gamma}{N^{\gamma+1}}\,\mathrm{d}N}.$$

For each credal function in $C_{N|r}$, we can find a prediction for N with a 95%-level confidence, by solving $c_\gamma(N \leq x|r) = 0.95$ for x, with $c_\gamma(N \leq x|r) = \int_r^x c_\gamma(N|r)\,\mathrm{d}N$.

We will find a prediction for N given by our imprecise posterior credal set $C_{N|r}$ by determining its upper bound, i.e., a prediction all distributions in $C_{N|r}$ can agree on. Now, as $\gamma \to 1$, the prediction for x such that $c_\gamma(N \leq x|r) = 95\%$ diverges. In other words, this imprecise representation of prior credence in N, reflecting our ignorance or indifference about N, doesn't yield any prediction about N.

Choosing any of the predictions given by the individual distributions in the credal set would be arbitrary. Without the possibility for my prior credence to be represented not by a single probability distribution but by an *infinite set* of probability distributions, I cannot avoid obtaining an arbitrarily precise prediction. Other distributions, such as distributions that decrease at different rates, could be added to the prior credal set, as long as they fulfill the criteria listed at the beginning of this section. However, no other distribution that we could include would change this conclusion.

3.3.3 Blurring Out Anthropic Predictions

We are ignorant about what value of the vacuum energy density ρ_V we should expect from our current theories. We can see that representing our prior ignorance or indifference about the value of the vacuum energy density ρ_V by an imprecise credal set can limit, if not entirely nullify, the role of anthropic considerations beyond that of mere boundary conditions.

If we substitute imprecise prior and posterior credences in the formula from (Weinberg, 2000, see infra § 4.3.3), we have $\mathrm{d}C_{\rho_V} = \nu(\rho_V) \cdot C^\star_{\rho_V}\, \mathrm{d}\rho_V$, with $C^\star_{\rho_V}$ a prior credal set that will exclude all values of ρ_V outside the anthropic range, and $\nu(\rho_V)$ the average number of galaxies which form for ρ_V, which as in § 4.3.3 peaks around the mean value of the anthropic range. In order for the prior credence $C^\star_{\rho_V}$ to express our ignorance or indifference, it should be such that it doesn't favor any value of ρ_V.

With the imprecise model, such a state of ignorance can be expressed by a set of probability distributions $\{c^\star_i : c^\star_i \in C^\star_{\rho_V}\}$, all of which normalizable over the anthropic range and such that $\forall \rho_V, \exists\, c^\star_i, c^\star_j \in C^\star_{\rho_V}$ such that ρ_V is favored by $c^\star_i$ and not by $c^\star_j$.[14] Such a prior credal set will not favor any value of ρ_V. In particular, it is in principle possible to define this prior credal set so that for any value of ρ_V, the lowest expectation (with respect to our credence) among the posteriors is lower than the highest expectation among the priors. If then we adopt interval dominance as a criterion for comparative confidence claims (see infra § 3.3.1), then no observation of ρ_V will be able to lend support to our anthropic prediction.

One may object to the adoption of interval dominance in such a case. This criterion is arguably not fined-grained enough to help us for most of the inferences we are likely to encounter. However, this choice of demanding confidence comparison rule can be motivated by the fact that we have no plausible alternative theoretical framework to the anthropic argument. In this context, it can be reasonable to agree to increase one's credence about the anthropic explanation *only if* it does better than any other yet unknown alternative might have done. Nonetheless, if we adopt other confidence comparison rules, it is possible with the imprecise model to construct prior credal sets that define a large interval over the anthropic range such that the confirmatory boost

[14] This can be obtained, for instance, by a family of Dirichlet distributions (preferable in order to have invariance under redescription (see de Cooman et al., 2009)), each of which giving an expected value at a different point in the anthropically allowed range. As in § 3.3.2, a lower bound can be placed on the variance of all the functions in $C^\star_{\rho_V}$ in order to avoid dogmatic functions.

obtained after observing ρ_V is not nearly as vindicative as it is with a single, uniform distribution.

This approach doesn't prevent Bayesian induction altogether. Because all the functions in $C^{\star}_{\rho_V}$ are probability distributions, they all can be updated as in usual Bayesian inferences and, in principle, converge toward a sharper credence, provided sufficient updating.

3.4 Conclusion

These cosmic puzzles show that, in the absence of an adequate representation of ignorance or indifference, a logic of induction will inevitably yield unwarranted results. Our usual methods of Bayesian induction are ill-equipped to allow us to address either puzzle. I have shown that the imprecise credence framework allows us to treat both arguments in a way that avoids their undesirable conclusions. The imprecise model rests on Bayesian methods, but it is expressively richer than the usual Bayesian approach that only deals with single probability distributions.

Philosophical discussions about the value of the imprecise model usually center around the difficulty to define updating rules that don't contradict general principles of conditionalization (especially the problem of dilation). But the ability to solve such paradoxes of confirmation and avoid unwarranted conclusions should be considered as a crucial feature of the imprecise model and count in its favor.

Bibliography

Bostrom, N. (2002). *Anthropic Bias: Observation Selection Effects in Science and Philosophy*. Routledge.

Carroll, S. M. (2001). The Cosmological Constant. *Living Reviews in Relativity*, 3(1):1–56.

Dawid, A. P., Stone, M., and Zidek, J. V. (1973). Marginalization Paradoxes in Bayesian and Structural Inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, 35(2):189–233.

de Cooman, G. and Miranda, E. (2007). Symmetry of Models versus Models of Symmetry. In Harper, W. L. and Wheeler, G., editors, *Probability and Inference. Essays in Honour of Henry E. Kyburg Jr*, pages 67–149. College Publications, London.

de Cooman, G., Miranda, E., and Quaeghebeur, E. (2009). Representation Insensitivity in Immediate Prediction under Exchangeability. *International Journal of Approximate Reasoning*, 50(2):204–216.

Dieks, D. (1992). Doomsday–Or: The Dangers of Statistics. *The Philosophical Quarterly*, 42(166):78–84.

Dieks, D. (2007). Reasoning about the Future: Doom and Beauty. *Synthese*, 156(3):427–439.

Goodman, S. N. (1994). Future Prospects Discussed. *Nature*, 368(March):106–107.

Gott, J. R. (1993). Implications of the Copernican Principle for our Future Prospects. *Nature*, 363(6427):315–319.

Gott, J. R. (1994). Future Prospects Discussed. *Nature*, 368(March):108.

Griffiths, T. L. and Tenenbaum, J. B. (2006). Optimal Predictions in Everyday Cognition. *Psychological Science*, 17(9):767–773.

Hartle, J. B. and Srednicki, M. (2007). Are We Typical? *Physical Review D*, 75(12):123523.

Huntley, N., Hable, R., and Troffaes, M. C. (2014). Decision Making. In Augustin, T., Coolen, F., and Troffaes, M. C., editors, *Introduction to Imprecise Probabilities*, pages 190–206. Wiley & Sons, Chichester, U.K.

Joyce, J. M. (2010). A Defense of Imprecise Credences in Inference and Decision Making. *Philosophical Perspectives*, 24(1):281–323.

Kyburg, E. H. (1978). Subjective Probability: Criticisms, Reflections, and Problems. *Journal of Philosophical Logic*, 7(1):157–180.

Leslie, J. A. (1990). Is the End of the World Nigh? *The Philosophical Quarterly*, 40(158):65–72.

Maor, I., Krauss, L., and Starkman, G. (2008). Anthropic Arguments and the Cosmological Constant, with and without the Assumption of Typicality. *Physical Review Letters*, 100(4):041301.

Martel, H., Shapiro, P. R., and Weinberg, S. (1998). Likely Values of the Cosmological Constant. *The Astrophysical Journal*, 492(1):29–40.

Monton, B. and Roush, S. (2001). Gott's Doomsday Argument. *http://philsci-archive.pitt.edu/id/eprint/1205*, pages 1–23.

Neal, R. M. (2006). Puzzles of Anthropic Reasoning Resolved Using Full Non-indexical Conditioning. *arXiv:math/0608592*, pages 1–56.

Norton, J. D. (2010). Cosmic Confusions: Not Supporting versus Supporting Not. *Philosophy of Science*, 77(4):501–523.

Pogosian, L., Vilenkin, A., and Tegmark, M. (2004). Anthropic Predictions for Vacuum Energy and Neutrino Masses. *Journal of Cosmology and Astroparticle Physics*, 7(005):1–17.

Richmond, A. (2006). The Doomsday Argument. *Philosophical Books*, 47(2):129–142.

Smith, C. A. (1961). Consistency in Statistical Inference and Decision. *Journal of the Royal Statistical Society. Series B (Methodological)*, 23(1):1–37.

Solà, J. (2013). Cosmological Constant and Vacuum Energy: Old and New Ideas. *Journal of Physics: Conference Series*, 453(1):012015.

Syversveen, A. R. (1998). Noninformative Bayesian Priors. Interpretation and Problems with Construction and Applications.

Tegmark, M. and Bostrom, N. (2005). Is a Doomsday Catastrophe Likely? *Nature*, 438(7069):754.

Troffaes, M. C. (2007). Decision Making under Uncertainty Using Imprecise Probabilities. *International Journal of Approximate Reasoning*, 45(1):17–29.

Vilenkin, A. (1995). Predictions from Quantum Cosmology. *Physical Review Letters*, 74(6):4–7.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London.

Weinberg, S. (1987). Anthropic Bound on the Cosmological Constant. *Physical Review Letters*, 59(22):2607–2610.

Weinberg, S. (2000). A Priori Probability Distribution of the Cosmological Constant. *arXiv preprint astro-ph/0002387*, pages 0–15.

Weinberg, S. (2007). Living in the Multiverse. In Carr, B., editor, *Universe or Multiverse?*, chapter 2, pages 29–42. Cambridge University Press.

Wells, W. (2009). *Apocalypse When? Calculating How Long the Human Race Will Survive*. Springer Praxis Books. Praxis.

Chapter 4

Lost in the Multiverse: Self-locating uncertainty, typicality, and observation bias

4.1 Introduction

This morning I was feeling terrible. The flu, maybe. Or maybe it was just that I'm not a morning person. I couldn't tell where I was. I remembered I'd been in an airplane recently, maybe I was traveling and that's why I felt lost. I could have been home or at a friend's home. Or maybe I was visiting my family in France. I was so foggy that I couldn't see or hear anything that would help me locate myself. I knew I would have to get out of bed and go get a cup of coffee: in what language would I greet the first person I would meet? I'm home in America much more often than I am in France; should I have assumed that I was probably at home? But when I travel to Europe I visit many people and as a consequence I stay over in many more places than when I'm home; should I have assumed that I was equally likely to have awaken in any of these rooms, and therefore that I was probably traveling? Maybe this was all a dream and, after all, maybe I should have assumed that I was more likely to be where most people are: somewhere in Asia between the latitude of Mumbai and that of Manila.[1]

Now that I've had several cups of coffee, I'm ready to tackle the question of how

[1] Or in a galaxy far far away identical to ours.

self-locating uncertainties should affect my beliefs, and of when typicality assumptions are legitimate and helpful. These questions weren't really crucial for me this morning—I found a door and opened it, and then I went down a flight of stairs without falling down: I knew that I could only be home. But answers to what appear to be similar questions may in other contexts be very elusive and yet decisive.

There are in the cosmology literature several kinds of arguments referring to self-locating uncertainty. Eternal inflation predicts the existence of a "multiverse" containing an infinity of "pocket universes" with fundamental physical constants taking any and all possible values. We don't know how to locate ourselves in this ensemble of possible universes. We don't know how to locate ourselves within our own universe either. We don't know if we are typical observers and if we can safely assume that the physical laws we draw from our observations hold elsewhere or if on the contrary we are very atypical observers, i.e., observers whose view on the world is particularly biased spatially or temporally.

A kind of self-locating uncertainty that is sometimes invoked in such a discussion comes from the possible existence of *copies of myself*. For instance, if the multiverse theory is true, then there are other pocket universes where reside observers having experiences indistinguishable from my own. As a consequence, we should be uncertain as to which of these pocket universes we inhabit. If one of these pocket universes is more favorable to the advent of life than the other (because of a difference in the value of some physical parameter), then this pocket universe is more likely to host a greater number of such copies of myself. This adds to the uncertainty about my location: I don't know which of these pocket universes I may find myself in, and, within this pocket universe, I don't know which of the copies of myself I am.

This kind of uncertainty is used for instance in anthropic reasoning, where predictions are obtained by conditionalizing a prior probability distribution for a given cosmological parameter on the number of observers it allows for (see, e.g., Weinberg, 1987; Bousso, 2006). Moreover, according to, e.g., Max Tegmark, there is even a real possibility that, in a universe that is large enough, "your closest identical copy is 10 to the 10^{28} meters away," (Tegmark, 2003, 48); i.e., within that distance, each of us has a doppelgänger with *exactly* the same experiences and memories.[2]

There are two main issues concerning all the above questions and arguments that require clarification. One is the status of self-locating beliefs and their role in induction and confirmation. The first question I want to address is how distinct are the following "two sorts of uncertainty": one "about what the world is like" and another one "about one's own spatial and or temporal location in the world" (Elga, 2000, 143). Should the realization that I *could* very possibly be somewhere else affect my other, non-indexical beliefs about the world? And if so, does that mean that my self-locating uncertainty should have a role in confirmation and induction distinct from other kinds of uncertainty?

A second, related question concerns *how* to handle self-locating uncertainty. In particular, I will ask whether and when assumptions of typicality (or a Copernican principle, or an assumption of mediocrity) are warranted or even required to characterize our self-locating uncertainty.

All these arguments involving our "location" broadly construed are not particularly new, whether in cosmology or in philosophy.[3] But the topic of self-locating beliefs (or,

[2] An argument made by, e.g., Page (2010); Aguirre and Tegmark (2011) and which I won't examine in more detail claims that this has problematic implications for Born's rule because, in a large universe, there may be other copies of ourselves that have exactly the same experiences but projection operators assume that there is only one observer. On this account, the only difference between us and copies of ourselves is our location, but which location each copy is in isn't available to them. Our expectation of projection operators (or other operators that would replace them) should be, Page claims, the average of what copies of ourselves could observe:

> Born's rule works when one knows where the observer is within the quantum state (e.g., in the quantum state of a single laboratory rather than of the universe), so that one has definite orthonormal projection operators. However, Born's rule does not work in a universe large enough that there may be identical copies of the observer at different locations, since then one does not know uniquely where the observer is or what the projection operators are. (Page, 2010, 2)

[3] For a critique of the anthropic principle, see, e.g., (Earman, 1987), for one of the Copernican principle, see, e.g., (Beisbart, 2010).

following Lewis (1979), "*de se* beliefs") and how they are related to *de re* beliefs has recently generated much discussion in philosophy. And arguments given as solutions to the Sleeping Beauty problem—a thought experiment introduced as an illustration of the distinctness of self-locating uncertainty and its bearing on our credence about the world—are still debated, in philosophy and cosmology alike.

In this paper I claim that the Sleeping Beauty problem can indeed help us find clarity about the nature and role of self-locating beliefs and uncertainty in confirmation, but not for the reasons usually given in the literature. I will show that the Sleeping Beauty problem doesn't involve anything pertaining specifically to self-locating beliefs. The canonical presentation of this problem complicates it unnecessarily: memory loss or self-locating uncertainty don't play any essential role in this problem. The Sleeping Beauty problem is in fact a problem about how to handle observation bias and as such does not constitute a new or distinctive challenge or source of rules for confirmation theory.

I will show that claims resting on typicality assumptions as to our location, on the other hand, represent a distinctive kind of claim. They rest questionable assumptions, such as the assumption that all models equally compatible with our observations are equally likely. Appealing to the Sleeping Beauty problem to justify this assumption rests on confusions about that problem.

4.2 Self-Locating Beliefs and Observation Bias

4.2.1 The Sleeping Beauty Problem as a Problem about Observation Bias

Are there, as Elga (2000) put it, "two sorts of uncertainty"—one "about what the world is like" and another one "about one's own spatial and or temporal location in the world"—that are distinct in such a way that they require different handling in confirmation theory or epistemology more broadly? The Sleeping Beauty problem was brought back to the forefront by Elga (2000) in order to illustrate that they are distinct, and to show how the former bear on the latter. I will contend that the kind of information and uncertainty the Sleeping Beauty problem purports to deal with is in fact not specific to self-location, and that as a consequence it can't be used to answer the questions it was intended to answer.

Here's the standard setup of the problem: Beauty will be put to sleep for three days on Sunday night. Right after she falls asleep, a fair coin will be tossed to determine how many times she will be briefly woken. If the coin toss results in *Heads*, Beauty will be briefly woken only once, on Monday. If *Tails*, she will be woken twice: once on Monday, and once on Tuesday. But after each waking, Beauty will be put back to sleep with a drug that makes her completely forget about that waking. Now, if we ask Beauty, right after she has been woken but without telling her what day it is, what her credence should be that the coin came up *Heads* on Sunday night, what should her answer be?

Two answers are usually given to this question: "$\frac{1}{2}$" and "$\frac{1}{3}$". For "halfers" (e.g., Lewis, 2001; White, 2006), that credence doesn't depend on self-locating beliefs, and at no point did she learn anything that would alter her credence that the coin is fair and that there is a one in two chance that it landed *Heads* on Sunday night. For "thirders" (e.g., Elga, 2000; Dieks, 2007; Titelbaum, 2008), upon waking up in the middle of the experiment, Beauty's credence should change merely because she finds herself in a different situation than before the experiment. Without having any new information, her self-locating uncertainty makes her waking after *Tails* twice as likely as after *Heads*.

More precisely, according to Elga (2000), and with H: "the coin landed Heads," T: "the coin landed Tails," M: "it is now Monday," U: "it is now Tuesday," Beauty believes the coin to be fair and therefore should give equal credence to $T\&M$ and $H\&M$. Moreover, because she's not able to locate herself, she should give equal credence to $T\&M$ and $T\&U$. Because $H\&M$, $T\&M$, and $T\&U$ are the only possible predicaments

Beauty could find herself in upon awaking, Elga asserted that the respective probabilities to be located in one of them sum up to 1. Consequently, Beauty's credence that the coin landed *Heads* on Sunday if asked in the middle of the experiment should be $\frac{1}{3}$. I will come back to Elga's argument and its relevance for anthropic arguments later in § 4.3.3.

Thus we can see what role self-locating uncertainty should play for our beliefs about what the world is like according to thirders. I contend, however, that this isn't what this problem illustrates. The canonical presentation of the Sleeping Beauty complicates it unnecessarily and obfuscates its meaning.[4] Contrary to claims made by, e.g., Titelbaum (2013a, §9), when Beauty awakes in the middle of the experiment, she has not *lost* certainty about her *location* as much as she has *gained knowledge* about her newly acquired *observation bias*. In the middle of the experiment, she knows that she will be awoken and asked about her credence in H twice as often after a T toss than after a H toss. Outside of the experiment, she may assume that her questioner's behavior won't depend on the outcome of the coin toss. But by entering the experiment, she acquires information about her questioner's asymmetric behavior, rather than losing certainty about her temporal location.

The Sleeping Beauty problem can indeed be reformulated as one about observation bias, in which self-locating uncertainty or memory loss play no role. Consider that, instead of having to use fictitious memory-loss-inducing sleeping pills, Beauty takes part in the following experiment, being fully aware of its setting: a quizzer is sitting at a table behind a screen, with his head above the screen; he regularly throws a fair coin, which then falls completely silently on a shock-absorbing mat; he asks Beauty what side she thinks the coin last landed, but he asks that question twice in a row without throwing the coin again each time it lands *Tails*. The problem now is to know what Beauty should answer in the middle of that experiment. Like in the standard version of the problem, she never knows if the question she answers follows a *Heads* toss or a *Tails* toss.

[4]Moreover, one might argue that doing so by invoking gratuitously a woman being asked to take memory-erasing pills is inconsiderate.

Beauty will on average achieve a better ratio of correct answers (and even a perfect one in an ideal situation) if she chooses to answer "Heads" only a third of the time. Therefore, if she wants to increase her chance of giving a correct answer to the quizzer, Beauty should indeed modify her answers when she is in the middle of the experiment. She shouldn't do so because she doesn't know where she is as much as because she has been made aware of the bias (observer bias or sampling bias) of the quizzer for whom *Tails* tosses count twice as much.

In the standard problem and this alternate experiment alike, if she didn't know of that bias, and if she were able to keep a tally of her correct guesses, Beauty would conclude that if the coin is fair, then the quizzer is biased (or vice versa). She could for instance start to suspect that the quizzer is cheating or can't see half the *Heads* results. But if Beauty is cognizant of her questioner's bias and if in the long run she concludes that one third of the answers should be "*Heads*", this would corroborate her belief that the coin is fair.

This problem, however, is nothing special for confirmation theory. Observation bias in any measurement is handled in a similar manner. However we choose to present the Sleeping Beauty problem, neither memories nor location—let alone self-locating uncertainty—play a distinctive role such that it requires a special handling, distinct from usual evidential reasoning.[5]

4.2.2 Beauty's Bets in a Rigged Game

Bradley and Leitgeb (2006); Cisewski et al. (2015), following earlier work by Seidenfeld et al. (1990), have argued that the Sleeping Beauty problem is an example where her credence and her betting behavior don't match. Neither memories, uncertainty nor indifference about one's location need be assumed in this betting framework either.

According to de Finetti (1974), the probability that an agent assigns to an event E

[5] Claims that self-locating uncertainty plays no such role can also be found in, e.g., (Bradley, 2011).

can be elicited by asking how much she is willing to bet that E, knowing that she would earn \$1 if E and \$0 otherwise. That price, p, is the *elicited probability* of E. If p is the price of the gamble and x the probability of E, the expected utility of the gamble is $x - p$. A fair bet is one where the agent expects neither gain nor loss ($p = x$).

Consider the alternative version of the Sleeping Beauty problem previously introduced, with the questioner hidden behind a screen and regularly asking Beauty how much she is willing to bet that the last coin toss was *Heads*, but asking her twice as often after each *Tails* toss. In this situation, for Beauty's bet to be fair to her and the bookie (i.e., the questioner), she has to account for the fact that in the long run, she will lose her wager twice as often as she will receive \$1. If she believes the coin to be fair (i.e., $x = 2$), the only way that neither she nor the bookie wins in the long run is if her wager is $\$\frac{1}{3}$. See Appendix 4.A for details.

In either version of the Sleeping Beauty problem (i.e., the standard version or the one with a coin tossed behind a screen and where no sleep is involved), her betting price is equivalent to her credence in *Heads* before observation bias is corrected for. In other words, her betting price is the elicited probability of a biased event, i.e., her credence about a skewed sample.

4.3 Self-Locating Uncertainty and Typicality in Cosmology

4.3.1 Typicality Assumptions

If self-locating uncertainty had a role to play in finding the value of fundamental parameters in our universe and throughout the multiverse, then it would be, as Titelbaum (2013b) argued, a good reason to care about the Sleeping Beauty problem indeed. There are in the cosmology literature several kinds of arguments referring to self-locating uncertainty. I will show that arguments appealing to the Sleeping Beauty problem to handle self-locating uncertainty are ill-founded. I will show that legitimate worries about our location, and worries about whether it should bear on our credence about what the world is like, can and should be handled as any other information about our observation bias. We will see that, as a consequence, typicality assumptions about our location or our existence as observers are unwarranted.

Typicality arguments are used to make predictions in the multiverse, predictions about, e.g., the value of the vacuum energy density. The most straightforward way to combine quantum field theory and general relativity leads to a dramatically incorrect estimate—usually cited as off by 120 orders of magnitude![6] A proposed solution to this problem is to predict this value from anthropic considerations (see, e.g., Weinberg, 1987; Bousso, 2006). The idea is to conditionalize the distribution for this parameter on the number of observers it allows for. Bousso and Freivogel summarized this as follows:

> We would like to predict low energy physics parameters observers are likely to observe. This requires statistical sampling of the theory landscape; an understanding of how the cosmological dynamics favors or disfavors the production of each vacuum; and finally, a sensible method for estimating the abundance of observers in each vacuum. (For example, parameters unique to a vacuum with no observers have zero probability of being observed.) (Bousso and Freivogel, 2007, 2)

This anthropic approach was first put forth by Weinberg (1987); the structure of his argument is as follows:

1. We first determine what range of values for that parameter are theoretically allowed. We first identify upper and lower bounds, outside of which the expansion of the universe is either too fast or too slow to allow for the advent of life. In the absence of

[6] See, e.g., (Zee, 2010)

a reason to think that the probability distribution would change over the relevant, anthropic scale, a uniform probability distribution within that range is assumed to be the least committal.[7]

2. We estimate the number of observers as a function of that parameter value.

3. We conditionalize the probability to observe different values of that parameter on the number of observers they allow, assuming that we are typical observers.

Later observations of the cosmological constant have been taken to vindicate this reasoning (see Weinberg, 2007). The value of the cosmological constant and other cosmological parameters is now much more constrained, but only observationally. Cosmologists nowadays still appeal to similar arguments in order to test the probative value of different theoretical scenarios. On this account, we place ourselves in a situation where we "forget" the known data about these parameters and assess how well different theories can "predict" them.

The result of such predictions rests on a choice of prior probability distribution (in step 1). However, with eternal inflation, the number of "universes" (pocket universes in the multiverse) can be infinite, and so can be the number of observers created. It is assumed that a prior probability distribution would be based on a measure over the space of "universes". The question of the choice of an appropriate measure and its definability is known as the measure problem in cosmology. Even though it cannot be derived from fundamental physics, such a measure is often assumed to exist.[8] In the absence of relevant fundamental theories, choices of prior probability distributions are often justified by appealing to the principle of indifference (or principle of insufficient reason), according to which all different parameter values within some allowed range should be considered

[7] I have discussed this assumption in greater detail in Chapter 3 above.

[8] See, e.g., (Linde and Noorbala, 2010) for a review of different proposed measures. All measures so far suffer from different severe problems, such as the youngness paradox (whereby it is extremely likely that we find ourselves as close as possible to the big bang), or the Boltzmann brains problem (whereby our data are extremely likely to be the product a random thermal fluctuation).

to be equally likely unless we have a good reason to think otherwise. This symmetry of belief, however, is entirely epistemic, not physical. This probabilistic interpretation of indifference is notoriously contentious and leads to notorious paradoxes of probabilistic inference.[9] Yet it is commonly used in cosmology, so much so that a uniform prior probability distribution over cosmological models in the range $-1 \leq \Omega_\kappa \leq 1$ (where Ω_κ characterizes the spatial curvature of the universe) is referred to as "the Astronomer's prior" (see, e.g., Trotta, 2012, § 11.3.1.1).

Parallel to the measure problem, the nature and role of typicality has been much discussed among cosmologists. This discussion involves separate, distinct—and some problematic—claims about our typicality as observers. But we can see that arguments used to make predictions based on our alledged typicality as observers sometimes conflate the following two types of typicality assumption:

- *Typicality assumption #1 (typicality of our data with respect to purely indexical information)*: In a large universe, other observers may have experiences indistinguishable from ours (see, e.g., Tegmark, 2003). If our doppelgängers and we perform a measurement of a same physical process or event, then we should expect that we won't all make the same observation (i.e., our measurements will differ). According to Bousso et al., we should, in such a case, act "as if our laboratory either [were] the only laboratory in the universe or [were] selected at random from among all the laboratories doing the same experiment in the universe. This is the assumption of typicality." (Bousso et al., 2008, 1)

- *Typicality assumption #2 (anthropic reasoning)*: In the multiverse, a "pocket universe" whose parameters are more favorable to the advent of observers is more likely to be observed.[10]

[9] See above Chapters 2 and 3.

[10] In a sense, this is a non-problematic tautology: a universe more favorable to the advent of life is more likely to be osberved by living entities. But in the literature this is meant to imply that such a universe will be more likely to be observed *by us*.

> In the multiverse, we can use typicality to make statistical predictions for the results of observations. For instance, to predict the cosmological constant, we would first determine the theoretically allowed values, and then count the number of observations of each value. The probability to observe a given value of the cosmological constant is proportional to the number of observations, in the multiverse, of that value. (Bousso et al., 2008, 2, continuation of the previous quote)

The first kind of typicality assumption is simply the assumption that predictions differing only on indexicality should be considered as equally likely. In this first kind of assumption, our prediction doesn't depend on our "location as observer": if we think that a stochastic phenomenon has a $\alpha\%$ chance of being observed given conditions x, y and z, then we should consider that any observer under conditions x, y and z has an $\alpha\%$ chance of observing this phenomenon. Such a prediction isn't affected by what can be considered as purely indexical information in the context of this experiment (e.g., the observer's name, her age, the color of her shirt, whether the experiment is conducted here or in an exact duplicate of our "pocket universe", etc.).

For example, imagine that I want to assess whether a coin factory is producing fair coins. To do so, I can ask many testers to take a coin each and run tests on it, and then average their results. Alternatively, I can take a coin at random and run tests on it myself. In doing so, I assume that the coin is typical of what the factory makes. Bousso et al. give a similar example to illustrate what they mean by an assumption of typicality, which I call 'typicality assumption #1.'

The second kind of typicality assumption is the assumption that we should conditionalize a theoretical prediction on the number of "observations" allowed by our theory. On this account, assuming that we are a "typical" observer means that, other things being equal, our observations are most likely to be those shared by the greatest number of observers. But if the content of the observations is determined by the number of observers, then it is the same anthropic argument as in step #3 of (Weinberg, 1987) given above.

Discussions of typicality in the cosmology literature sometimes simply assert that "our observation should be taken at random among all possible observations." But this formulation is ambiguous between the two distinct ideas noted above. It can be understood either as asserting typicality in the first sense. In that case, it asserts that the probability of an observation only depends on the physical process observed, not on who is observing or on how many observers are performing the same observation. Or it can be understood as asserting typicality in the second sense. In that case, however, the probability of an observation will depend on the total number of observations (in the sense given above), which greatly depends on how we are defined as observers. For instance, in their daily life, what academics typically see (printed words or a computer screen) differs significantly from what sentient life forms typically see (prey, rocks, and plants, I assume). Presumably, academics fall into both categories of observers, and therefore, if we adopt this second kind of typicality assumption, predictions about what a typical observer sees will depend on an arbitrary choice of reference class of observers.

In order to see how my predictions and inferences would differ depending on which version of typicality assumption we adopt, consider my situation when I woke up this morning, having no idea where I was. According to typicality assumption #1, if upon opening my bedroom door I saw an American-shaped electric outlet, then, other things equal, I should have assumed that I was in America. But according to typicality assumption #2, I should have assumed, *a priori*, that I was most likely somewhere in India,[11] and the evidence provided by the outlet might not be enough to outweigh the factor due to the population weighting.

In the literature, however, the two kinds of typicality assumption are not always carefully distinguished. This has led to confusing exchanges. According to Hartle and Srednicki, the claim that we shouldn't prefer a theory that would make us typical ob-

[11] Or in a galaxy far far away

servers to one that wouldn't is "absurd", "contrary to the standard scientific practice of assigning noninformative priors that do not preselect conclusions" unless we have good reasons to do so, and a "mere personal preference for theories in which we are typical of something." (Hartle and Srednicki, 2007) In response to such criticism, Bousso and Freivogel invoked the first kind of typicality assumption to justify the broader, ambiguous claim about typicality, arguing that "the overall success of the scientific method so far suggests that [the typicality assumption] is appropriate." (Bousso et al., 2008, 1)

One can find subsequent reformulations of this broader typicality assumption in terms of our "location" in the multiverse, such as the following: "In determining where in the multiverse we are living, we make the assumption of typicality: *we are equally likely to be anywhere consistent with our data.* This is called the 'principle of indifference'." (Freivogel, 2011, 4, emphasis mine) If we consider that every region $\upsilon_1, \ldots, \upsilon_n$ of the multiverse susceptible of hosting life (that is, equally compatible with our data) is equally likely to be realized (as we did earlier in step 1 of Weinberg's argument), and if we then count the number of "locations" that can possibly be occupied by observers in each of these regions (i.e., the number of observers) $\nu_1, \ldots, \nu_n$, then we should, on this account, consider that we are equally likely to find ourselves in one of these $L_n = \sum_{i=1}^{n} \upsilon_i \cdot \nu_i$ locations, each with a probability of $\frac{1}{L_n}$, which makes it a priori more probable that I find myself in a region of the multiverse where the greatest number of observers are.[12]

We will first see, in § 4.3.2, that attempts to justify this line of argument based on solutions to the Sleeping Beauty problem rest on confusions about the nature of that problem. We will then see, in § 4.3.3, that the motivation behind conditionalizing on the number of observers initiated by Weinberg and still invoked today is analogous to an argument made to defend the thirder's position about the Sleeping Beauty problem; and I will claim that this argument is flawed.

[12]Here, Freivogel echoes a method of "observation averaging" suggested in (Page, 2010): "I advocate first constructing the ensemble of probabilities [i.e., the probabilities of life-permitting regions of the multiverse] and then using typicality within that wider ensemble."(Freivogel, 2011, 4, n. 1)

4.3.2 Confusions about the Sleeping Beauty Problem

Interestingly, we can sometimes find in the cosmology literature explicit references to the Sleeping Beauty problem (and the thirder solution in particular) to justify the broader, ambigious formulation of typicality assumption according to which, as we just saw in § 4.3, we should consider that we are equally likely to be in the multiverse in any of the locations consistent with our data; that is, that we are equally likely to be any of the observers having experiences indistinguishable from our own.

If Beauty adopts this typicality assumption about her location, then, in the standard setting of the Sleeping Beauty problem, she will conclude that she is equally likely to wake up in one of the three possible locations: on Monday after a *Heads* toss, on Monday after a *Tails* toss, or on Tuesday, necessarily after a *Tails* toss. However, this claim rests on confusions about the Sleeping Beauty problem. Indeed, this claim entails the following interpretation of the Sleeping Beauty problem, which for convenience I will call

the Indifferent Sleeping Beauty problem:

1. before she is put to sleep, Beauty is told that there is a nonzero probability that she will either be woken up once or twice,
2. she will have no memory of any possible awakening until the experiment is over, and
3. each of the possible awakenings is as likely as any other.

With this interpretation of the Sleeping Beauty problem the probability of *Heads* is consistent with the standard thirder answer to the Sleeping Beauty problem. However, unlike in the standard version of the problem (or the one I gave, which is equivalent), the probability of *Heads* is given to Beauty as a premise. In the Sleeping Beauty problem, Beauty's equal credence in her awaking in one of the three possible predicaments (Monday after Heads, Monday after Tails, or Tuesday) is due to the fairness of the coin. In the standard Sleeping Beauty problem, there's no equivalent to point #3 in the Indifferent Sleeping Beauty problem.

Predictions in the multiverse that assume that we are typical observers based on the assumption that, based on an entirely epistemic principle of indifference, all scenarios generating our locations are equally likely are analogous to the Indifferent Sleeping Beauty problem. These predictions are akin to a version of the problem in which we tell Beauty that she should be indifferent about what determines her location when she wakes up. In such a version of the Sleeping Beauty problem, she doesn't need to know what process will determine whether she will be awoken once or twice and can just assume that whatever that process, it will make her location typical. In terms of the betting framework seen earlier in § 4.2.2 (and in Appendix 4.A), it's equivalent to asking one's wager without telling them what the bet is about.

It's easy to see how such predictions don't follow from the standard version of the Sleeping Beauty problem. Suppose that, instead of a coin toss, it is the throw of a fair die that will decide how many times she will be awoken: once if the die comes up 1, twice otherwise. There is now only $\frac{1}{6}$ chance that she will only be awoken once.

We can arrive at this result by correcting for Beauty's questioner's bias: she's twice as likely to be asked a question after a number other than 1 was rolled, which is five times more likely to occur than the roll of a 1. We have $P(2 \text{ to } 5) = 10 \times P(1)$ and $P(1) + P(2 \text{ to } 5) = 1$, and therefore $P(1) = \frac{1}{11}$. This result is incompatible with point #3 in the Indifferent Sleeping Beauty problem.

Assuming that our place among observers in the multiverse is typical comes to assuming that the initial conditions for each "pocket universe" (at least those that are life-permitting) are equally likely. With that assumption, making predictions in the multiverse would rest solely on how favorable to life each "pocket universe" is; it wouldn't depend anymore on the prior probabilities for their initial conditions. This indeed would amount to asserting, without any justification, what Hartle and Srednicki (2007) call a "mere personal preference for theories in which we are typical of something." This is closer to a form of strong anthropic principle than to usual scientific practice.[13] We will see in § 4.3.3 in greater detail what argument could be more rigorously based on the thirder solution to the Sleeping Beauty problem, but we will also see why it is unwarranted.

It should be noted that if one uses the work of Adam Elga—and in particular his defense of the principle of indifference in the case of self-locating uncertainty in (Elga, 2004)—to justify assumptions of typicality in the multiverse, one in fact makes "an absurd claim that [Elga] do[es]n't endorse," namely that all physical processes or hypotheses having the same observable consequences are equally likely.[14] Elga distinguished this claim from another, uncontroversial claim that he endorses, namely that states differing only on indexicality deserve equal credence (which doesn't mean that our credence should necessarily be evenly *divided* among all states differing only on indexicality). Sebens and Carroll (2015) discussed a roughly formulated principle of indifference for cases of self-locating uncertainty adapted from (Elga, 2004), according to which "an observer should give equal credence to any one of a discrete set of locations in the universe that are consistent with the data she has"; to avoid possible confusion about this claim, we should add the following proviso, implicit in (Elga, 2004): "provided all those locations are equally likely to exist."[15]

[13] Bostrom (2002) has pointed out what he calls the problem of the Presumptuous Philosopher: if the likeliest theory is, *a priori*, the one that predicts the largest number of observers, then any theory choice could be made by any presumptuous armchair philosopher.

[14] Or with Elga's terminology: "centered worlds representing indistinguishable predicaments deserve equal credence" (Elga, 2004, 387).

[15] To be sure, conditional on the multiverse scenario, all possible observers are guaranteed to exist. The relative likelihood of their existence is determined in one region of the multiverse or another by the ratio of the occurrence of these regions. There, Sebens and Carroll are concerned with claims evoked earlier in fn. 2 according to which indifference in cases of self-locating uncertainty would amount to branch counting—and therefore have disastrous empirical consequences—in Everettian quantum mechanics. However, it is questionable whether considerations about self-locating uncertainty add anything new in this context, or to our usual treatment of predictions in physics more generally.

4.3.3 Anthropic Reasoning

Now, let's go back to the argument presented at the beginning of this section in § 4.3.1 and assume that we have a well-defined, physically-motivated prior distribution over the possible values of a given cosmological parameter. What should we make of step 3 of that argument, namely conditionalizing a prior distribution over the possible values for the cosmological constant on the number of observers each of them allows for?

We saw that, according to, e.g., Weinberg (1987); Bousso (2006), we should, other things being equal, have greater credence in a cosmological model that is more favorable to the advent of life, because then it is more likely that it will be observed. There are different possible ways to do that. Conditionalizing a distribution for a given parameter on some of its observable consequences is uncontroversial. One might think that by taking the *number of galaxies* as a proxy for the number of observers, Weinberg conditionalized his predictions on an observable consequence. Indeed, we can determine the number or the density of galaxies in our observable universe. But that's not what he suggested: by *only* conditionalizing on the number of galaxies as a function of a given parameter value, he assumed that we are more likely to find ourselves in a universe with the greatest possible number of galaxies.

Sometimes challenges to this sort of anthropic argument focus on whether the number of galaxies is the right proxy (see, e.g., Aguirre, 2001), or point out that its results depend heavily on the choice of reference class, that is, on how "observer" is defined: does any life form count, only intelligent life forms, or should the amount of "observation time" matter more than the number of observers? But such challenges don't always question the assumption that I can take my own existence (as an observer) to be an indication of anything more than the mere fact that I exist (for instance, that it is *likely* that I exist).[16]

[16] A related assumption is what Bostrom calls the "Self-Indicating Assumption," according to which "[g]iven the fact that you exist, you should (other things equal) favor hypotheses according to which many observers exist over hypotheses on which few observers exist." (Bostrom, 2002, 66). As we already

Anthropic reasoning proceeds from the assumption that we are typical observers among all possible observers, which was typicality assumption #2 in § 4.3.1. In the cosmology literature, this assumption is usually presented as an intuition, but not justified. We can see, however, that, if the thirder solution Elga (2000) gave as a solution to the Sleeping Beauty problem is valid, then it can be generalized to support anthropic reasoning. I will present Elga's solution in greater detail than in § 4.2.1, then formulate a plausible generalization of this argument. I will then show that Elga's argument is flawed, and so is the generalized version.

Recall the quantities and equalities invoked by Elga (2000). They can be summarized in the following table:

	Monday	Tuesday
Heads	$cr(H\&M) = A$	$\varnothing$
Tails	$cr(T\&M) = B$	$cr(T\&U) = C$

According to Elga, we have:

$$\begin{cases} A = B \\ B = C \\ A = B = C = \dfrac{1}{3} \end{cases} \tag{4.1}$$

More precisely, to make this assertion, Elga first argued that, conditional on being awake on Monday (which occurs whether the coin lands *Heads* or *Tails*), your credence in H is just the chance P_H that H is true.[17] Here, one must assume that it doesn't make a difference to our credence in H whether it is determined before or after the first room

saw in Chapter 3 and above in n. 13, Bostrom doesn't endorse this assumption (see Bostrom, 2002, §7).

[17]Elga has to assume that our credence in H, $cr(H)$, can differ from the chance that H occurs, P_H. Otherwise the question asked to Beauty about her credence in H in the middle of the experiment wouldn't be interesting (it would be one half whether or not she is in the middle of the experiment as long as she believes the coin to be fair).

has already been created (i.e., whether the coin is thrown on Sunday night or on Monday night).[18] We then have:

$$cr(H|M) = cr(T|M) = P_H. \tag{4.2}$$

Second, Elga claims that, conditional on T, my credences are evenly divided among the two times Beauty could be awake (Monday or Tuesday).[19] That is, we have

$$cr(M|T) = cr(U|T). \tag{4.3}$$

Those two conditions and Bayes theorem determine the ratio $\dfrac{cr(H)}{cr(T)}$ as follows:

$$\frac{cr(H)}{cr(T)} = \frac{cr(M|T)}{cr(M|H)} \cdot \frac{cr(H|M)}{cr(T|M)} = \frac{\frac{1}{2}}{1} \cdot \frac{P_H}{P_T} = \frac{1}{2}, \tag{4.4}$$

since it is assumed that $P_H = P_T$.

The generalization of this solution to an argument in support of anthropic reasoning follows immediately from the special case.[20]

Imagine that, for all $k \in \{1 \dots N\}$, the number of people created n_k is determined by the throw of a $N-$sided die (not necessarily a fair die). Let h_k be the proposition that k was rolled. Let P_k be the chance that h_k is true. Assume also that, for all k, if h_k is true then n_k identical rooms are created, each occupied by one of the n_k people. Let r_i be the proposition that I am in room i. The generalized Sleeping Beauty problem is the following question: knowing only, for all hypotheses h_k, the corresponding values P_k and

[18]To illustrate the intuitive motivation for this argument, consider the following scenario: you have a flat tire in the middle of the woods and you call Otto+, your auto insurance. But you don't remember what insurance policy you purchased: the normal plan, with which they send you a mechanic, or the "above and beyond" plan that not only guarantees you a mechanic but also, five minutes later, a chamber orchestra to provide you high-quality entertainment while your tire is being changed. The arrival of the mechanic doesn't change either of your credences that you purchased one policy or the other, since she will come whichever policy you purchased.

[19]From the setup of the problem, conditional on H, my credence in being woken on Monday in the middle of the experiment must be 1.

[20]I am indebted to Wayne Myrvold for this generalized argument's formulation.

n_k, and that I am in one of the n_k rooms, what should be my credence in h_k?[21]

We can summarize the situation with the following table, for two such hypotheses h_i and h_j:

	r_1	...	r_{n_i}	...	r_{n_j}
h_i	$cr(h_i \& r_1)$	...	$cr(h_i \& r_{n_i})$	$\varnothing$	$\varnothing$
h_j	$cr(h_j \& r_1)$	...	$cr(h_j \& r_{n_i})$	...	$cr(h_j \& r_{n_j})$

We can therefore consider that my situation as an observer taking part in that experiment is analogous to that of Sleeping Beauty: it is as if I were asked about my credence in h_k and knew that, conditional on h_k, I could equally likely be placed in any of the r_{n_k} rooms.

The argument proceeds in the same two steps as in (Elga, 2000). First, we need to argue that, conditional on being in the first room (which is occupied on all hypotheses), my credence in h_k is just the chance that h_k is true. To do so, one must assume, like we did earlier, that it doesn't make a difference to our credence in h_k whether it is determined before or after the first room has already been created (i.e., whether the die is rolled before or after the creation of the first room). In order to make this hypothesis, we have to consider that, like in the Sleeping Beauty problem, when I am asked about my credence in h_k, I have no memory of several "awakenings" that may have occurred in the past. In the context of anthropic reasoning, we can give the following analogy: if h_k is true, then rooms in which there are observers who have experiences identical to my own are created at different times. For this situation to be analogous to the Sleeping Beauty problem, we can assume that the complete generation of all possible observers happens in two times: a first observer is created (it might be me, for all I know), and then all other possible observers will be created at once (I might be one of them, and I don't know how many of them there are). How many other observers will be created

[21]The Sleeping Beauty problem is the special case where $N = 2$, h_1: *Heads*, h_2: *Tails*, $P_1 = P_2$, and $n_2 = 2 \cdot n_1 = 2$.

depends on h_k, the chance of which we know to be P_K. Like in the Sleeping Beauty problem, someone with experiences identical to mine existed before h_k was decided, and I don't know if I find myself before or after this moment.[22] Then, on this account, for all hypotheses, conditional on being the first observer ever created, my credence in h_k should just be the chance of h_k, namely P_k.

In other words, with this assumption, we have a general formulation of the first conditions given in the special case—i.e., eq. (4.2):

$$\forall k,\ cr(h_k|r_1) = P_k. \tag{4.5}$$

Second, we show that for all k and conditional on h_k, my credences are evenly divided among $r_1, \ldots, r_{n_k}$. That is, we have the following generalization of eq. (4.3):

$$\forall k, \forall i, j \in \{1 \ldots n_k\},\ cr(r_i|h_k) = cr(r_j|h_k) = \frac{1}{n_k}. \tag{4.6}$$

Those two conditions and Bayes theorem determine, for any i, j, the ratio $\frac{cr(h_i)}{cr(h_j)}$ the following generalization of eq. (4.4):

$$\frac{cr(h_i)}{cr(h_j)} = \frac{cr(r_1|h_j)}{cr(r_1|h_i)} \cdot \frac{cr(h_i|r_1)}{cr(h_j|r_1)} = \frac{n_i}{n_j} \cdot \frac{P_i}{P_j}. \tag{4.7}$$

According to this argument, it is therefore rational to set my credences in h_i and h_j such that $cr(h_i)/cr(h_j)$ is boosted by n_i/n_j, which is exactly what anthropic reasoning dictates we should do.

Cisweski and her collaborators, in (Cisewski et al., 2015, § 3), have claimed that "thirder" position , assumed in this generalized argument, rests on an inadequate partition of Beauty's space of possibilities. According to them, the flaw lies in the fact

[22] In the Sleeping Beauty problem, Beauty doesn't know if she is awoken on Monday or Tuesday.

that the sets of events $\{M\&H, M\&T, U\&T\}$ or $\{M, U\}$ don't form a partition of her space of possibilities. Yet Elga's inductive probabilistic reasoning by case (whereby he determines an unconditional credence from conditional credences) requires that we reason from Beauty's exhaustive and mutually exclusive possibilities (i.e., a partition of her space of possibilities).

Consequently, Cisewski et al. argue, if the set of events from which we reason is exhaustive but is not a partition, then probabilistic reasoning follows what they call the "Law of Too Much Probability". For example, for two conditions χ_1 and χ_2 such that $\chi_1 \cup \chi_2 = \Omega$ and $\chi_1 \cap \chi_2 = \chi_3 \neq \varnothing$, then we have

$$cr(E) = cr(E|\chi_1) \cdot cr(\chi_1) + cr(E|\chi_2) \cdot cr(\chi_2) - cr(E|\chi_3) \cdot cr(\chi_3). \tag{4.8}$$

In the middle of the experiment, Beauty can't have access to any information that would help her know when "today is Monday" is true and "today is Tuesday" is not. She is unable to substitute either specific day for "now", and, for her in the middle of the experiment, "now" can't pick out anything other than $M \vee U$. Therefore, on this account, the partition of Beauty's space of possibilities can only be $\{H\&M, T\&(M \vee U)\}$, which is equivalent to simply $\{H, T\}$. Thus Beauty's credence in H shouldn't change in the middle of the experiment.

If one grants that the partition in the Sleeping Beauty problem is a flawed one, then one has to conclude that Elga's generalized argument, too, suffers from an inadequate partition. My earlier presentation of this argument didn't distinguish carefully between two situations: being located in time *before* or *after* the creation of the first room. Without this distinction, the analogy with the Sleeping Beauty problem would break down, and then we couldn't make the assumption formulated in eq. (4.5) that, conditional on being in the first room, my credence in h_k is just the chance P_k that h_k is true.

In order for the generalized argument to resemble more closely and unequivocally the Sleeping Beauty problem, we should make the distinction between the following two situations:

- r_1^*: "I am in room 1 *before* the roll of the die that determines whether h_k is true," and

- r_1: "I am in room 1 *after* the roll of the die that determines whether h_k is true,"

so that r_1^* is analogous to M, and $r_1, \ldots, r_{n_k}$ to T.

With this additional distinction, we make sure that the set $\{r_1^*, r_1, \ldots, r_{n_k}\}$ is exhaustive of all the situations an observer could find herself in. But, like Beauty during the experiment, this distinction wouldn't be apparent to that observer. According to Cisewski and collaborators, this distinction can't affect this observer's partition, and the anthropic argument fails.

However, one might argue that the issue of whether Beauty can see the set of events $\{M\&H, M\&T, U\&T\}$ as a partition doesn't depend on whether she can distinguish its members. Yet one need not agree with Cisewski and her collaborators to see why conclusions from the Sleeping Beauty problem don't carry over to anthropic reasoning.

In effect, the conditions for observation bias to occur and affect Beauty's credence about the coin toss don't exist in the anthropic case. In both cases, our credence in a physical event depends on the chance of that event (which depends on the coin's fairness in the Sleeping Beauty problem, and on the relative presence of a part of the multiverse having properties favorable to the advent of life in the case of anthropic reasoning). However, the distinction made above between the two "locations" r_1^* and r_1, necessary for the anthropic argument to be analogous to the Sleeping Beauty problem, would be completely artificial. Indeed, there is nothing in the anthropic argument considered above that plays a role analogous to that of Beauty's questioner. That is, in the multiverse, we are only "awake" once; we only occupy one location in the multiverse. Consequently, in the anthropic argument, there is no room for the kind of observation bias that is at play in Elga's problem.

Therefore, eq. (4.5) should hold only for r_1^* (i.e., before the die roll), but not for r_1 (i.e., after the die roll), and for clarity we should replace it with the following:

$$\forall k,\ cr(h_k|r_1^*) = P_k. \tag{4.2'}$$

Moreover, eq. (4.6) should only apply to $r_1, \ldots, r_{n_k}$ (i.e., *after* the die roll), and we have

$$cr(h_k) = \sum_{i=1}^{n_k} cr(h_k|r_i) \cdot cr(r_i), \tag{4.9}$$

which is just the law of total probability if we forget the artificial situation r_1^*. In eq. (4.9), eq. (4.2') is irrelevant. Consequently, unlike with eq. (4.4), there is nothing we can deduce as to the relative likelihood of two hypotheses h_i, h_j based on the relative number of observers they produce. Thus eq. (4.4) becomes

$$\frac{cr(h_i)}{cr(h_j)} = \frac{P_i}{P_j}. \tag{4.4'}$$

Knowing that I exist and without observing any other humans, this doesn't allow me to assert that I should have more credence in a hypothesis based only on the fact that it is more favorable to the advent of life. On the other hand, with the partition of events $\{r_1, \ldots, r_{n_k}\}$ (i.e., without considering r_1^*), we can use Bayes theorem to have

$$\forall k,\ cr(h_k|r_1) = \frac{P_k}{n_k \cdot cr(r_1)}. \tag{4.2'}$$

Although this prevents anthropic considerations to alter our credence in any hypothesis h_k, eq. (4.2') allows us to update our credence in the hypotheses h_k upon learning r_1, which eq. (4.5) didn't allow for. Indeed, eq. (4.5) assumed, without any physical justification, that all hypotheses conditional on being in the first room were equally likely. As we saw, without the questionable assumption in eq. (4.5), knowing that I exist and without observing any other humans, I *shouldn't* have more credence in a hypothesis based only on the fact that it is more favorable to the advent of life.

In § 4.3.1, we saw that considerations about how typical a hypothesis makes us shouldn't affect our credence about that hypothesis. Appendix 4.B shows how to reach the same conclusion within the betting framework. In this section, we saw how anthropic reasoning could be made analogous to the Sleeping Beauty problem. Since the Sleeping Beauty problem is an exemplary puzzle about how self-locating beliefs should affect our credences (much like anthropic considerations should do if anthropic reasoning were valid), one might have hoped that, if the thirder position were true, it could be used to justify anthropic reasoning. We saw, however, that the new element the "thirder" solution purports to bring to such arguments is flawed. Consequently, we are compelled to conclude, as we did earlier in § 4.3.1, that favoring hypotheses on the grounds that they make our location more typical is unwarranted.

Therefore, in the case of the cosmological constant problem for instance, the mere fact that we exist and how favorable to the advent of life a model is cannot guide our cosmological predictions and confirmations, whether or not we have a physically well-motivated prior probability distribution over the different possible value of the vacuum energy density. We can, however, conditionalize on observable consequences of the possible values a parameter may take. But that has nothing to do with anthropic reasoning.

4.3.4 Testing the Copernican Principle

Previously we wondered whether we should have greater credence in a cosmological model if it allows for the existence of more observers having experiences indistinguishable from ours. It is possible to rephrase that question as follows: should we have greater credence in a cosmological model if it allows for the existence of more locations from where observations would be similar to our own? If so, then if we have a choice between two models compatible with our data, then, other things being equal, we should have greater credence in one that doesn't require us to occupy a special position. This is what the Copernican principle prescribes, and it has very real implications for our choice of cosmological models.

Consider a cosmological model characterized by the usual six free parameters of the ΛCDM model describing the matter-energy content of the universe, the spatial distribution of primordial density fluctuations, and the effect of ionizing radiation.[23] One of its parameters corresponds to the density of baryonic (i.e., non-dark) matter, and it is not given by background theoretical assumptions. Therefore, it can only be constrained from observation.[24] To do so, we assume that, at large scales, the distribution of matter is homogeneous and isotropic. This assumption is called the Cosmological Principle, and it plays a transcendental role in much of cosmology: the homogeneity and isotropy of the universe at large scales are essential ingredients for the Friedmann-Lemaître-Robertson-Walker metric, an exact solution of Einstein Field Equations of general relativity considered to be the standard model of cosmology.[25]

Such cosmological models as well as the interpretation of our cosmological data usually rest on the assumption that the apparent spatial homogeneity and isotropy doesn't stem from our having a very special point of view on our cosmic neighborhood. In doing so, we assume that we are not atypical observers, i.e., that the distribution of matter we can observe is representative of the average distribution of matter in the universe. To justify this last assumption, one can invoke the principle of sufficient reason: we shouldn't presume that there is anything special about our cosmic surroundings, unless we have a good reason to do so.

[23]For the latest data see, e.g., (Planck Collaboration, 2014). Such a model characterizes our "pocket universe". I am not here considering the possible existence of other "pocket universes".

[24]Such constraints on matter density or the cosmological constant can be data from the cosmic microwave background, baryonic acoustic oscillations, and measurements from supernovae (see, e.g., Gong et al., 2013).

[25]The isotropy of the universe at large scale is an empirically established fact, but only for the observable universe. This fact is used to support the adoption of the Copernican principle, which may be empirically testable. However, the isotropy and homogeneity at even larger scales—beyond the boundaries of the observable universe—is what the Cosmological principle asserts. This assumption isn't testable to the same extent as the Copernican principle. Yet the Cosmological principle is necessary to derive the FLRW metric. On the distinction between these two principles and their testability, see, e.g., (Beisbart and Jung, 2006).

Adopting the Copernican principle has consequences for our choice of cosmological models:

> This implies that the spacetime metric reduces to a single function of the cosmic time, the scale factor $a(t)$. (...) Low redshift observations combined with the assumption of almost flatness of the spatial sections, justified mainly by the cosmic microwave background data, lead to the conclusion that (...) the expansion is accelerating. This conclusion involves no hypothesis about the theory of gravity or the matter content of the universe, as long as the Copernican principle holds. This has stimulated a growing interest in possible explanations, ranging from new matter fields dominating the dynamics at late times to modifications of general relativity. (Uzan et al., 2008, 1)

In other words, if we endorse the Copernican principle, we need to account for the apparent accelerating expansion of the universe. As suggested in the last quote, important aspects of our theoretical background are at stake (see Uzan, 2010; Huterer, 2011, for reviews). We are now facing a choice: we can either hold on to the Copernican Principle but then may have to postulate a mysterious dark energy, or we can renounce the typicality of our location. As Uzan et al. argued, renouncing the Copernican Principle would have as a consequence that "we may be living close to the center (because isotropy around us seems well established observationally) of a large underdense region." (Uzan et al., 2008, 1)

Let $cr(H)$ denote our credence that the universe around us is isotropic and homogeneous (which is what the Copernican Principle asserts), $cr(\neg H)$ our credence that it is isotropic but inhomogeneous. Let $\{r_n\}$ be the set of n possible locations compatible with our data about our cosmic surroundings. Like we had in § 4.3.3, there will be far many more such locations with H than with $\neg H$. Now, let's ask, like in § 4.3.3, what role, if any, should the fact that H makes our location and the observed isotropy typical play in our credence in H? How confident should we be that our location is typical even if we have a good reason to suspect that it isn't (such as the the need to posit dark energy)?

This situation is analogous to anthropic reasoning discussed in § 4.3.3. We can indeed use eq. (4.7) to compare our respective credence in H and $\neg H$ relative to how typical our location is under each hypothesis:

$$\frac{cr(H)}{cr(\neg H)} = \frac{cr(H|r_H)}{cr(\neg H|r_{\neg H})} \cdot \frac{cr(r_{\neg H}|\neg H)}{cr(r_H|H)}, \tag{4.10}$$

with $r_H, r_{\neg H}$ the total number of locations from which we could observe isotropy under hypotheses $H, \neg H$, respectively. The conclusions from §§ 4.3.1, 4.3.3 apply here as well: as I have shown in Appendix 4.B, the betting framework compels us to assert that considerations about our typicality and the number of locations compatible with our observations don't affect our prior credences $cr(H)$ and $cr(\neg H)$. I have shown that attempts to argue otherwise don't succeed: either such attempts conflate two senses of typicality assumption, or they rest on flawed probabilistic arguments. Consequently, the answer to questions about the role of self-locating uncertainty for testing the Copernican principle will be identical to that given to questions about the role of anthropic considerations.

Now, we can see that for this inquiry, like for the other problems we considered previously, obtaining additional evidence—and not considering relative self-locating uncertainties as we did it—can guide us. In this case, for instance, Uzan et al. (2008) showed that measuring the time drift of cosmological redshifts would provide us with that sort of evidence. Consider our past light cone (see Fig. 4.1, left). We move on our universe line, and so does a galaxy, observed at a time t with redshift z. Consider then our past light cone at a time $t + \delta t$. That same galaxy has also moved during δt and is now seen with a redshift $z + \delta z$ (see Fig. 4.1, right).

The redshift is linked to the scale factor $a(t)$ as follows:

$$1 + z = \frac{\lambda_{rec}}{\lambda_{em}} = \frac{a_0}{a}, \tag{4.11}$$

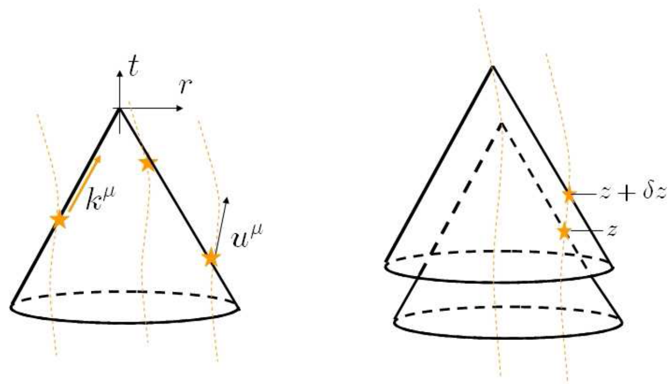

Figure 4.1: Illustration of the drift δz over δt, from (Uzan, 2010, § 1.2.4)

where a_0 is the scale factor at the time of the emission, and a that at the reception. If going from z to $z + \delta z$ is compatible with the Copernican principle, then we have:

$$\dot{z} = H_0(1+z) - H(z), \tag{4.12}$$

with H_0 at emission and H at reception (see Sandage, 1962) Now, if only isotropy is assumed, *but not homogeneity*, then for an observer at the center of a spherically symmetric universe, we have:

$$\dot{z} = H_0(1+z) - H(z) + \frac{1}{\sqrt{3}}\sigma(z), \tag{4.13}$$

with $\sigma(z) = 0$ if homogeneous. z can be determined observationally (e.g., with type Ia supernovæ) without assumptions regarding homogeneity.

Assuming we can carry out this experiment,[26] considerations about the relative degree of our self-locating uncertainty will not help us settle this issue. Either we observe an inhomogeneity (i.e., $\sigma(z) \neq 0$, which would confirm H) or we don't. If we observe evidence of the inhomogeneity of our cosmic neighborhood and at the same time remain committed to the Cosmological principle, then the adoption of the Copernican principle is unwarranted, regardless of considerations of the initial self-locating uncertainty borne by

[26] According to (Uzan, 2010, § 1.3.1), a typical order of magnitude is, for $z = 4$: $\delta z \approx -5 \cdot 10^{-10}$ over a period of time $\delta t \approx 10yr$. So-called "extremely large telescopes," now under construction, may allow for such observations.

different hypotheses about this homogeneity. In other words, what matters in an inquiry about our location (namely here, what our cosmic neighborhood is like) is empirical evidence about our location, not how typical each hypothesis about our location makes us.

4.4 Conclusion

There is a clear motivation to assess how self-locating uncertainty can bear on our knowledge of the world in cosmology where fundamental physical theories to explain, e.g., the value of the cosmological constant are lacking. Therefore, it's not surprising that the Sleeping Beauty problem has made its way through the cosmology literature. This in turn provides philosophers with an interesting playground. I have argued, on the contrary, that the Sleeping Beauty problem has nothing new to contribute to confirmation and evidential reasoning in cosmology; it merely dictates that one should modify their belief when they obtain information about their observation bias.

Typicality assumptions often used in cosmology intend to handle self-location uncertainty by assigning equal probability to locations or to models consistent with our data. Although this reasoning would yield the same solution to the Sleeping Beauty problem as one reasoning in terms of sampling bias, it rests on a principle of indifference with which our results are epistemic rather than physical and evidence-based in nature.

In this paper I have dealt with two different situations, and two corresponding typicality assumptions:

- In § 4.2.1, we saw that the Sleeping Beauty problem should be construed as a problem about how to correct for observation bias (i.e., how to handle information about, e.g., the reliability of our observation device). In this context, a typicality assumption tells us that, unlike information about our data or their reliability, information or uncertainty of purely indexical character shouldn't affect how we reason about the evidence we have. For instance, if Beauty didn't know about her questioner's bias, and if she could keep a tally of her correct responses (and if she considers her questioner's identity to be irrelevant in this regard), she would soon conclude that the coin is biased toward *Tails*.

- In § 4.3, in the case of anthropic reasoning or the Copernican principle, an assumption of typicality has the effect of giving a confirmatory boost to theories that make our position as observer (whether the number of observers or the number of observational standpoints) typical.

These two typicality assumptions have very different effects on confirmation: the former tells us that, without a good reason to suspect otherwise, we should treat our data as if they were representative of what anyone else in a similar situation (with respect to what we think are relevant conditions) would obtain; the latter tells us, without any physical justification, to have a bias towards theories that make us typical. In the context of anthropic reasoning, this would be equivalent to the claim that that nature is good to us; put differently, this second kind of typicality assumption favors cosmological models that are fine-tuned for life. Other things being equal, one type of assumption, the first one, is *neutral* with respect to confirmation and induction, whereas the second introduces an unwarranted bias.

Self-locating beliefs (i.e., our knowledge about our "location", broadly construed) can contribute to our knowledge of the world, but only qua knowledge of possible observation bias. As a consequence, it's not the case that self-locating uncertainty is distinct from uncertainty about what the world is like in a way that requires any special treatment in confirmation theory.

4.A Appendix A: Beauty's Credence and Betting Behavior

Below is a summary, adapted from (Cisewski et al., 2015), of how to calculate Beauty's credence about the coin toss's result based on de Finetti's method for eliciting a personal probability from betting behavior.[27]

According to de Finetti (1974), the probability that an agent assigns to an event E can be elicited by asking how much she is willing to bet that E, knowing that she would earn \$1 if E and \$0 otherwise. That price, p, is the *elicited probability of* E. If p is the price of the gamble and x the probability of E, the expected utility of the gamble is $x-p$. A fair bet is one where the agent expects neither gain nor loss ($p = x$).

More generally, for any variable X (a real-valued function defined on the state space Ω), p_X the agent's price for X, β_{X,p_X} a real number set up by the opponent, the gambles the agent will be willing to enter are of the form

$$\beta_{X,p_X}[X - p_X]. \tag{4.14}$$

For the agent, the bet is fair if $p_X = X$. If $\beta > 0$, the agent pays βp_X in order to receive βX in return, and if $\beta < 0$, the agent receives βp_X in order to pay βX in return. In any case, a gamble is advantageous to the agent if $X > p_X$ and more or less risky depending on the value of β.[28]

The general formula for the net cash flow of the series of bets Beauty can enter is as follows:

$$\sum_i \beta_i F(\omega)[C_i(\omega)(x_i - p_i)], \tag{4.15}$$

[27] In Appendix 4.B, I use this formalism in the context of many possible self-locations (e.g., when there are copies of myself).

[28] In the general case, X, p_X, β can be any real number. In the previous example, we had $x = X$, $\beta = \$1$, and $p = p_X$, with $x, p \in [0, 1]$. Having $\beta = 1$ makes p comparable to a probability measure. This is de Finetti's method for eliciting a personal probability (or credence).

with $i \in \{MH, MT, UT\}$, C the result of the coin toss (either $C = H$ or $C = T$), $x_i \in \{a, b, c\}$, and $p_i = p_{a,b,c}$ the price for the series of bets Beauty can enter. Let's assume that for all i, $\beta_i = \beta$. The net cash flow corresponding to the series of bets Beauty can enter is as follows:

$$\sum_i \beta_i F(\omega)[C_i(\omega)(x_i - p_i)] = \beta F(\omega)[H(\omega)(a - p_{a,b,c}) + T(\omega)(b - p_{a,b,c}) + T(\omega)(c - p_{a,b,c})] \tag{4.16}$$

$$= \beta F(\omega)[H(\omega)(a - p_{a,b,c}) + T(\omega)(b + c - 2p_{a,b,c})].$$

Beauty would pay $p_{a,b,c}$ twice if the coin lands *Tails*, but only once if *Heads*. Now, since for all ω, $H(\omega) = 1 - T(\omega)$, the net cash flow can be written as

$$\beta F(\omega)[H(\omega)(a - b - c + p_{a,b,c}) + b + c - 2p_{a,b,c}]. \tag{4.17}$$

The standard Sleeping Beauty problem (the first one presented above in §1) corresponds to $a = 1$ and $b = c = 0$. That is, upon awakening, Beauty is asked her price for the bet that MH happens (and therefore a is expected) and neither MT nor UT happens. The net cash flow then can be written as

$$\beta F(\omega)[H(\omega)(a - b - c + p_{a,b,c}) + b + c - 2p_{a,b,c}] = \beta F(\omega)[H(\omega)(1 + p_{1,0,0}) - 2p_{1,0,0}] \tag{4.18}$$

$$= \beta(1 + p_{1,0,0})F(\omega)\left[H(\omega) - \frac{2p_{1,0,0}}{1 + p_{1,0,0}}\right].$$

Since β can be determined after p has been announced,[29] we can rewrite the net cash

[29]Side note: we saw earlier that the sign of β would affect how to interpret the gambling process. But here $p_{1,0,0} > 0$ and therefore the sign of β doesn't depend on it.

flow as

$$\beta F(\omega)[H(\omega) - x], \tag{4.19}$$

with $x = \dfrac{2p_{1,0,0}}{1 + p_{1,0,0}}$. In this form, one can elicit Beauty's credence in $H(\omega)$. A fair price for this gamble is indeed such that $x = H(\omega)$, which corresponds to the probability of H in the event that $F(\omega) = 1$—i.e., the probability of H conditional on F, $P(H|F)$. This is a crucial result for the analysis of the role of self-locating uncertainty: a credence in an event and a fair price for a bet about that event are distinct, and they don't always have the same value. Now, assuming that Beauty is necessarily awake when she is asked about $P(H)$, ω is necessarily such that $P(F) = 1$ and $P(H|F) = P(H)$. Assuming that the coin is fair, we then have

$$P(H|F) = P(H) = \frac{1}{2}. \tag{4.20}$$

What Beauty considers to be a fair price for the standard Sleeping Beauty bet is not directly her credence in H. Indeed, we have $x = \dfrac{2p_{1,0,0}}{1 + p_{1,0,0}}$ and then

$$p_{1,0,0} = \frac{x}{2 - x} = \frac{1}{3}. \tag{4.21}$$

4.B Appendix B: Bets and Credence with Many Possible Self-Locations (or Copies of Myself)

We can see that, unlike the Sleeping Beauty problem, anthropic reasoning is not about observation bias. Within the betting framework, we can see that considerations about the number of other possible copies of myself I could be, or locations I could find myself in, won't affect my betting price.

Consider an example similar to that of § 4.3.3: a version of the Sleeping Beauty problem where if *Heads*, there will be m copies of myself, n if *Tails* ($m < n$ and as in § 4.3.3, assume that the first m in either scenario are identical). Knowing that the coin is fair, should that affect my betting price that the coin came up *Heads*?

Let's proceed as in Appendix 4.A. Beauty can enter one of $m + n$ possible bets (or that there are $m + n$ possible situations in which she can enter that bet): if H_1 (H and she's observer #1), if H_2, and so on. The payoff is a if H and b otherwise. Let's make the same assumption about β as in Appendix 4.A.

Following eq. (4.15), the net cash flow in this situation can be written as

$$\beta \left[\sum_{i=1}^{m} P(H_i)(a - p_{1,0}) + \sum_{j=1}^{n} P(T_j)(b - p_{1,0}) \right]. \tag{4.22}$$

Now, we have

$$\sum_{i=1}^{m} P(H_i) = P(H), \tag{4.23}$$

$$\sum_{j=1}^{n} P(T_j) = P(T) = 1 - P(H). \tag{4.24}$$

Unlike in the standard Sleeping Beauty problem, the price $p_{1,0}$ is paid only once if *Heads* and only once if *Tails*. Thus any consideration about self-locating uncertainty vanishes. Here, the betting price and credence are identical.

Bibliography

Aguirre, A. (2001). Cold Big-Bang Cosmology as a Counterexample to Several Anthropic Arguments. *Physical Review D*, 64(8):1–12.

Aguirre, A. and Tegmark, M. (2011). Born in an Infinite Universe: A Cosmological Interpretation of Quantum Mechanics. *Physical Review D*, 84(10):105002.

Beisbart, C. (2010). Can We Justifiably Assume the Cosmological Principle in Order to Break Model Underdetermination in Cosmology? *Journal for General Philosophy of Science*, 40(2):175–205.

Beisbart, C. and Jung, T. (2006). Privileged, Typical, or Not Even That? – Our Place in the World According to The Copernican and The Cosmological Principles. *Journal for General Philosophy of Science*, 37(2):225–256.

Bostrom, N. (2002). *Anthropic Bias: Observation Selection Effects in Science and Philosophy*. Routledge.

Bousso, R. (2006). Holographic Probabilities in Eternal Inflation. *Physical Review Letters*, 97:191302 1–4.

Bousso, R. and Freivogel, B. (2007). A Paradox in the Global Description of the Multiverse. *Journal of High Energy Physics*, 06(18).

Bousso, R., Freivogel, B., and Yang, I.-S. (2008). Boltzmann Babies in the Proper Time Measure. *Physical Review D*, 77(10):103514.

Bradley, D. (2011). Self-Location Is No Problem for Conditionalization. *Synthese*, 182(3):393–411.

Bradley, D. and Leitgeb, H. (2006). When Betting Odds and Credences Come Apart: More Worries for Dutch Book Arguments. *Analysis*, 66(2):119–127.

Cisewski, J., Kadane, J. B., Schervish, M. J., Seidenfeld, T., and Stern, R. (2015). Eliciting Sleeping Beauty's Credences. *(unpublished)*, pages 1–24.

de Finetti, B. (1972). *Probability, Induction, and Statistics*. Wiley, New York.

de Finetti, B. (1974). *Theory of Probability*. Wiley, New York.

Dieks, D. (2007). Reasoning about the Future: Doom and Beauty. *Synthese*, 156(3):427–439.

Earman, J. (1987). The SAP Also Rises: A Critical Examination of the Anthropic Principle. *American Philosophical Quarterly*, 24(4):307–317.

Elga, A. (2000). Self-Locating Belief and the Sleeping Beauty Problem. *Analysis*, 60(2):143–147.

Elga, A. (2004). Defeating Dr. Evil with Self-Locating Belief. *Philosophy and Phenomenological Research*, LXIX(2):383–396.

Freivogel, B. (2011). Making Predictions in the Multiverse. *Classical and Quantum Gravity*, 28(204007):1–15.

Gong, Y., Gao, Q., and Zhu, Z.-H. (2013). The Effect of Different Observational Data on the Constraints of Cosmological Parameters. *Monthly Notices of the Royal Astronomical Society*, 430(4):3142–3154.

Hartle, J. B. and Srednicki, M. (2007). Are We Typical? *Physical Review D*, 75(12):123523.

Huterer, D. (2011). The Accelerating Universe. In Goodstein, D., editor, *Adventures in Cosmology*, volume 25, chapter 11, pages 321–353. World Scientific Publishing Company.

Lewis, D. (1979). Attitudes De Dicto and De Se. *The Philosophical Review*, 88(4):513–543.

Lewis, D. (2001). Sleeping Beauty: Reply to Elga. *Analysis*, 61(3):171–176.

Linde, A. and Noorbala, M. (2010). Measure Problem for Eternal and Non-Eternal Inflation. *Journal of Cosmology and Astroparticle Physics*, 2010(09):008–008.

Page, D. N. (2010). Born's Rule Is Insufficient in a Large Universe. *arXiv preprint arXiv:1003.2419*, pages 1–7.

Planck Collaboration (2014). Planck 2013 Results. XVI. Cosmological Parameters. *Astronomy & Astrophysics*, (Forth.):1–69.

Sandage, A. (1962). The Change of Redshift and Apparent Luminosity of Galaxies due to the Deceleration of Selected Expanding Universes. *The Astrophysical Journal*, 136:319.

Sebens, C. T. and Carroll, S. M. (2015). Self-Locating Uncertainty and the Origin of Probability in Everettian Quantum Mechanics. *The British Journal for the Philosophy of Science (forthcoming)*, pages 1–41.

Seidenfeld, T., Schervish, M. J., and Kadane, J. B. (1990). When Fair Betting Odds Are Not Degrees of Belief. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1:517–524.

Tegmark, M. (2003). Parallel Universes. *Scientific American*, 288(5):40–51.

Titelbaum, M. G. (2008). The Relevance of Self-Locating Beliefs. *Philosophical Review*, 117(4):555–606.

Titelbaum, M. G. (2013a). *Quitting Certainties: A Bayesian Framework Modeling Degrees of Belief*. Oxford University Press.

Titelbaum, M. G. (2013b). Ten Reasons to Care about the Sleeping Beauty Problem. *Philosophy Compass*, 8(11):1003–1017.

Trotta, R. (2012). Cosmological Bayesian Model Selection: Recent Advances and Open Challenges. In Feigelson, E. D. and Babu, G. J., editors, *Statistical Challenges in Modern Astronomy V*, volume 902 of *Lecture Notes in Statistics*, pages 127–140. Springer New York, New York, NY.

Uzan, J.-P. (2010). Dark Energy, Gravitation and the Copernican Principle. In Ruiz-Lapuente, P., editor, *Dark Energy: Observational and Theoretical Approaches*, chapter 1, pages 3–47. Cambridge University Press.

Uzan, J.-P., Clarkson, C., and Ellis, G. F. R. (2008). Time Drift of Cosmological Redshifts as a Test of the Copernican Principle. *Physical Review Letters*, 100(19):191303–1–4.

Weinberg, S. (1987). Anthropic Bound on the Cosmological Constant. *Physical Review Letters*, 59(22):2607–2610.

Weinberg, S. (2007). Living in the Multiverse. In Carr, B., editor, *Universe or Multiverse?*, chapter 2, pages 29–42. Cambridge University Press.

White, R. (2006). The Generalized Sleeping Beauty Problem: A Challenge for Thirders. *Analysis*, 66(290):114–119.

Zee, A. (2010). The Cosmological Constant Problem and the Cosmic Coincidence Problems. In *Quantum Field Theory in a Nutshell*, chapter VIII.2, pages 448–451. Princeton University Press.

Chapter 5

Simplicity and Unification in Cosmological Model Selection

5.1 Introduction

In the last two or three decades, cosmology came of age and left a time with "only $2\frac{1}{2}$ facts"[1] to enter a "precision era" defined by increasingly precise measurements of a number of cosmic parameters. Driven by a new wealth of data, cosmologists have then turned their attention to statistical and Bayesian model comparison methods.

Model selection in cosmology consists in identifying the relevant parameters necessary to characterize our universe, its composition and evolution, and the values of these parameters. Cosmological parameters can be the cosmological matter, baryon, or radiation densities, the Hubble parameter characterizing the spatial curvature of the universe, the cosmological constant Λ, etc.(see Liddle, 2004, for a review of candidate parameters for cosmological models). The Λ-cold dark matter (ΛCDM) model, for instance, is a widely used parametrisation of the Big Bang model based only on 6 of such parameters. Competing models may require a different number of such parameters, each tuned at different values; for instance, the ΛCDM model can be extended to include cosmological inflation or other processes to account for the evolution of the early universe.

[1]According to Peter Scheuer in 1963, "1) the sky is dark at night, 2) the galaxies are receding from each other as expected in a uniform expansion, and $2\frac{1}{2}$) the contents of the universe have probably changed as the universe grows older." (quoted in Longair, 1993, 160).

Among such competing cosmological models, however, the values of some of their parameters (their free parameters) may be left undetermined by our background theories, and can only be constrained observationally. It is in that context that cosmologists have recently started to appeal to Bayesian and statistical model selection methods (see, e.g., Trotta, 2008, 2012; Liddle, 2009; Hobson et al., 2010, for recent reviews). These model selection—or, rather, comparison—methods rely on likelihoodist and Bayesian criteria such as the Bayes factor (the ratio of models' Bayesian evidence), the Bayesian Information Criterion (BIC), or other information criteria such as the Akaike Information Criterion (AIC).[2]

Bayesian methods often need to make reference to prior probability distributions; however, in cosmology, this would require measures over the space of cosmological models that are well-defined, well-behaved, but also physically motivated. In addition to conceptual and technical difficulties to define such measures (see, e.g., Schiffrin and Wald, 2012), fundamental theories motivating cosmological parameters are lacking. In the absence of well-motivated prior probabilities for cosmological models, cosmologists appeal for instance to what is called the "Astronomer's prior," a uniform distribution over a given range of models among which we are indifferent a priori.

To compare two models $\mathcal{M}_1$ and $\mathcal{M}_2$ assumed to be equiprobable, the Bayes factor—the ratio of the models' evidence $\frac{p(\mathrm{data}|\mathcal{M}_1)}{p(\mathrm{data}|\mathcal{M}_2)}$—selects the model giving the best fit between data and model. However, in the absence of well-motivated prior distributions for cosmological parameters and models, and without a priori constraints on how complex our models can be, fit-to-data alone is a poor selection criterion. Indeed, some models will be able to achieve better fit-to-data simply in virtue of having more free parameters.

Therefore, because model selection criteria such as the BIC or the AIC tend to optimize the number of free parameters (i.e., introduce a parameter only if it significantly

[2]There exist other information criteria (see Spiegelhalter et al., 2002; Konishi and Kitagawa, 2008, for reviews). Among these other criteria, the Deviance Information Criterion (DIC) is sometimes used in cosmology (Liddle, 2007). The BIC and the AIC are the two main such criteria, on which many other criteria are based.

improves fit-to-data), they are seen as more refined and more discerning than methods that merely tend to maximize the likelihood of data given a model. Both in the cosmology literature and the philosophy literature, they are often introduced as acting as an Ockham's razor, i.e., as seeking ontological parsimony.

This account is prevalent in cosmological selection in cosmology. For instance, Kunz et al. (2006) or Martin et al. (2011) used a measure of Bayesian complexity, a correction of the Bayesian evidence which simply measures a model's goodness-of-fit, meant to reward a model with parameters that are fewer and defined over smaller ranges. Relatedly, Andrew Liddle suggested to "use the Akaike and Bayesian information criteria to carry out cosmological model selection, in order to determine the parameter set providing the preferred fit to the data" and so as to conclude, for instance, that spatially flat models are "statistically preferred" to closed models (Liddle, 2004, 49).

The rationale behind the use of Bayesian complexity, the AIC, or the BIC, rather than the Bayes factor, is often limited to claims about the need for a balance between fit-to-data and model complexity (in terms of number of its free parameters). Indeed, main introductory texts justify this approach by invoking Ockham (see, e.g., Trotta, 2008, 2012; Hobson et al., 2010, §§ 5–7).[3] Thus they echo claims about the importance of *parsimony* as a model selection criterion that are widespread in the statistics literature as well as in central philosophical publications on model selection (see especially Forster and Sober, 1994; Forster, 2000; Sober, 2008, § 1.7). This view rests on the notion that, because selecting models based only on their ability to fit the data may lead to *overfitting* the data, we should instead weigh each parameter against its informativeness and penalize those parameters that increase fit only marginally or not at all. In other words, model selection should aim at reaching a balance between fit and complexity because we should avoid overfitting. It is this claim that I here want to assess.

[3]Liddle didn't explicitly mention the medieval logician in (Liddle, 2004), where he gave more elaborate arguments to justify these criteria (see also Liddle, 2009). But we will see in § 5.3 why the specific context of cosmology makes their use problematic.

In § 5.2.1, we will first see, for model selection in general and in cosmology in particular, what motivates the search for criteria that go beyond fit-to-data that likelihoodist methods estimate. We will see why it is often argued that simplicity should guide model selection in that regard. Some philosophers have disputed the relevance of simplicity—an ambiguous and possibly arbitrary notion—as a criterion for model selection. John Norton (2012a,b), for instance, recently argued that an information criterion such as AIC is in fact the result not of considerations about simplicity, but rather of background assumptions about our space of models under test. Following such arguments, I will claim in § 5.2.2 that whether or not model selection information criteria can be useful in cosmological model selection does not depend on any notion of parsimony or model complexity, and I will argue that these information criteria are *not* after parsimony or after a balance between fit and complexity.

Independently of the question of the relevance of simplicity in model selection, we will see, in § 5.3, that the use of statistical methods in cosmological model selection faces severe obstacles that limit our ability, in principle, to avoid overfitting the data. Furthermore, we will see that, because universes only come in very small numbers, the relevance of certain statistical methods—the AIC in particular—is questionable.

However, I will suggest, in § 5.4, that another Bayesian model selection method can be meaningful and relevant in cosmology. I will argue that there is a sense in which *not parsimony but rather unification*, or lack of dispersion, can guide model selection. Wayne Myrvold (2003) has given a Bayesian account of a model's ability to unify different sorts of phenomena, regardless of the number of parameters at play. The measure of information used in this account is one between a set of phenomena and another, otherwise unrelated set of phenomena, given a certain model. This measure gives a formal character to the idea of consilience of inductions (Whewell, 1847), according to which "a consilience of inductions would occur when the values of certain parameters can be determined from two different sorts of phenomena, and the values determined from one class of phenomena agree with those determined from another." (Myrvold, 2003, 418) Parameter estimation plays a role in assessing the unifying power of a model, but on this account it is the improved relationship between parameters of different kinds that will provide support to a choice of model. The sort of lack of dispersion this measure rewards is the consolidation of the relationship between phenomena.

Thus after having argued that in cosmology usual model selection methods that aim at maximizing fit-to-data and predictive accuracy 1) do not seek simplicity, and 2) are not adequate methods in many contexts, I will claim that carrying out model selection by measuring consilience, on the other hand, can be pursued by Bayesian methods and overcome problems related to the limited sample of universes accessible to us.

5.2 Simplicity and Model Selection

5.2.1 Why Simplicity Might Matter

A basic ingredient in model selection in cosmology is the Bayesian evidence $p(D|\mathcal{M})$, i.e., the likelihood of the data D given a model $\mathcal{M}$, defined as follows:

$$p(D|\mathcal{M}) = \int p(D|\theta, \mathcal{M})\, p(\theta|\mathcal{M})\, \mathrm{d}\theta, \tag{5.1}$$

with θ a particular set of values for the free parameters of a given model. That is, the model likelihood is the integral over all the possible parameter values of the product of the likelihood of the parameters in a given model and the prior for the parameters in that model.

To compare two models $\mathcal{M}_1$ and $\mathcal{M}_2$ assumed to be equiprobable, the Bayes factor—the ratio of the models' evidence $\frac{p(\text{data}|\mathcal{M}_1)}{p(\text{data}|\mathcal{M}_2)}$—selects the model giving the best fit between data and model.

This method of model comparison assumes that we can assign prior probabilities to the parameters at play in a given model. We will see later, in § 5.3, the challenges that this assumption poses for many inferences using the Bayesian evidence in cosmology. Assuming that the number of parameters is fixed, likelihood-based model comparison methods such as this one can tell us how closely model and data fit together. This method can be used for parameter estimation, that is, to determine for each parameter the value that best fits the data.

Model selection in cosmology, however, aims at comparing models with different numbers of parameters. Because fundamental theories that could explain, for instance, the initial conditions in the early universe or the value of the cosmological constant are lacking, model selection is used in the literature to *determine how many parameters our cosmological models should have.* This role is explicit, for instance, in several applications of cosmological model selection (see, e.g., Liddle, 2004; Kunz et al., 2006; Szydlowski et al., 2015; Martin et al., 2011)

However, if in model selection we didn't have any constraints regarding how to choose our parameters, and if our preferences between models were only a function of their Bayesian evidence, we would likely favor models that overfit the data (i.e., *ad hoc* models). For any given data set, the best-fitting linear curve, defined by only two parameters—i.e., a best-fitting curve of the form $f(x) = ax + b$—will necessarily not be a better-fitting curve than any best-fitting polynomial curve of higher order—i.e., a best-fitting curve of the general form $f(x) = \sum_{i=0}^{n>1} a_i x^i$. Consider for instance the data set in Fig. 5.1: the higher the order of the best-fitting polynomial curve, the better the fit to the data.

However, a better-fitting curve isn't necessarily more physically plausible. If for instance the data set in Fig. 5.1 includes data points collected under different experimental conditions which have not been corrected for, then seeking the best fit, whatever the number of parameters required to do so, could move us *away* from the physically true fitting curve. In fact, "a curve that is maximally close to the data (because it passes exactly through all the data points) is probably not going to be maximally close to the

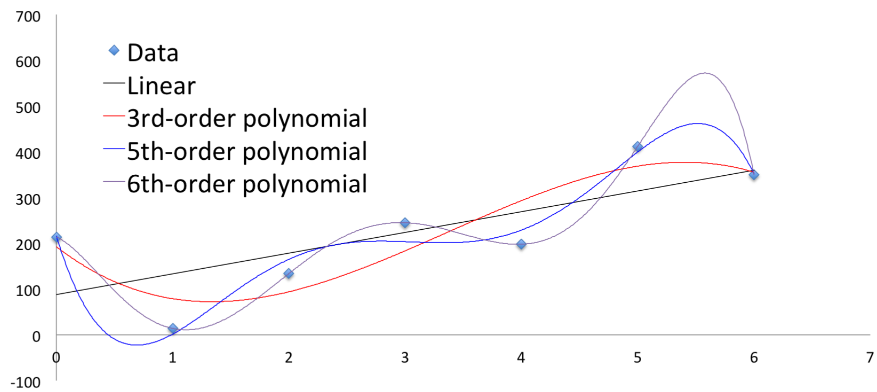

Figure 5.1: Different polynomial curves of different orders (from 1, linear, to 6). The higher the order, the better the curve fits the data; only the 6th-order curve fits all the data points.

truth. Closeness to the truth is different from closeness to the data." (Forster and Sober, 1994, 6) If we were simply interested in the best-fitting model for a given data set, then we should seek models that are maximally close to the data. But maximal accuracy to the data we already have is not necessarily a good indicator of further data we could gather in the future.

Imagine, for instance, that the data points in Fig. 5.1 are measurements of the temperature of a given amount of monoatomic gas at high temperatures and low pressures as the function of its volume. Our knowledge of thermodynamics—of the ideal gas law in particular—leads us to expect that it is a *linear* curve that will best represent the relationship between the two quantities measured here. However, it is only polynomials of order 6 and higher that will pass exactly through all the data points. Therefore, if we want to infer from this data set a model that is most likely to perform best, on average, with all such future measurements, then we should look for the *linear* curve that best fits the data we already have. Looking for a curve, a polynomial, of higher dimension would be looking for a model that *overfits* that data. That is, it would be looking for a model "which is too sensitive to idiosyncrasies in the data set that are unlikely to recur in further samples drawn from the same underlying distribution." (Hitchcock and Sober, 2004, 11) Indeed, it wouldn't make much physical sense to look for a model that characterizes a relationship described by Boyle's law by more parameters, which would likely be ontologically superfluous—i.e., ad hoc.

Therefore, although empirical adequacy is obviously a critical criterion in model selection, it cannot be the only one; we do not just want our models to fit our data (actual or potential), we want them to tell us what are the relevant parameters that determine them. And it is because likelihood-based methods may favor better-fitting yet unwarranted models that cosmologists, in need of additional guiding principles for induction, appeal to some sense of *parsimony*.

An illustration often given to illustrate the importance of parsimony and the danger of adhocness is the example of Ptolemaic versus Copernican astronomy (see, e.g. Forster and Sober, 1994, § 5); if fit-to-data and predictive accuracy were the only relevant criteria for model selection, then we should prefer the Ptolemaic model, because, unlike with the Copernican model, we can always add epicycles to improve it.

In order to avoid adhocness and overfit, the cosmology literature, echoing the philosophy and statistics literature, usually invokes Ockham's razor, i.e., the prescription not to "multiply entities unnecessarily":

> Bayesian model comparison makes use of an Occam's razor argument to rank models in term of their quality-of-fit *and* economy in the number of free parameters. A model with more free parameters will naturally fit the observations better, but it will also be penalized for the wasted parameter space that the larger number of parameters implies. (Kunz et al., 2006, 1)

By appealing to "Ockham's razor" as a principle that applies to the number of entities (or parameters), cosmologists follow a widespread but puzzling tradition of crediting the medieval logician with the maxim "*Entia non sunt multiplicanda præter necessitatem.*" However, as discussed in, e.g., (Thorburn, 1918), this formulation of the principle of parsimony cannot be found in the writings of Ockham himself. What one can find in the writings of Ockham are the following formulations: "pluralities ought not be supposed without necessity," and "in vain we do by many that which can be done by means of fewer."[4] However, as argued in, e.g., (Ariew, 1977), it is debatable to view Ockham as holding a metaphysical principle of parsimony with respect to the number of entities and the simplicity of Nature. Moreover, none of these versions of the principle of parsimony was specifically Ockham's; they were quite common among his contemporaries in one form or another, and this principle was commonly attributed by Scholastic thinkers to Aristotle and Aquinas.[5]

Such claims by cosmologists echo arguments found in, e.g., (Forster and Sober, 1994; Sober, 2008, § 1.7) claiming that statistical or Bayesian information criteria can serve as such an Ockham's razor by weighing a model's accuracy against its number of free parameters. In practice, for a given level of accuracy a model attains, such information criteria score that model according to the level of informativeness of its parameters. These information criteria thus "shave off" superfluous parameters, guiding us toward the most economical and (hopefully) relevant models, thereby reducing the risk of overfitting. On this account, it is because of this resistance to overfitting that a simpler model (simpler with respect to the number of its free parameters) should be preferred.

[4] "*pluralitas non est ponenda sine necessitate*" for the former, and "*frustra fit per plura quod potest fieri per pauciora*" for the latter. Although Trotta (2008, n. 3), for instance, gives the first of these two formulations, he interprets it in terms of number of free parameters, as is commonly done by cosmologists and philosophers of science.

[5] "One is forced to conclude that Ockham's razor is not Ockham's. Ockham was not the first to have coined "entities must not be multiplied without necessity"; he had no part in formulating it. Ockham was not the most avid user of principles of parsimony; the principle Ockham used to reduce the ontology of his realist opponents was his principle of absolute divine omnipotence, a principle of possible plenitude. Ockham did hold methodological principles of parsimony, but he was not the first to coin these, either. Ockham must have regarded his principles as methodological and must have been careful not to state them as a metaphysical doctrine. Ockham's views on metaphysics and theology seem to have been inconsistent with his holding a metaphysical principle of parsimony." (Ariew, 1977, 17)

5.2.2 Why It Doesn't, and What Model Selection Criteria Require

There is no doubt that overfitting should be avoided. But assuming something to the effect that "the simplest explanation is always the best"[6] seems at odds with scientific practice, which is more interested in finding not the simplest but the *truest* explanation.

The oft-repeated maxim attributed to Ockham, however, isn't itself simple. What exactly does it mean for a theory to be simpler or to have fewer entities? And why would nature care about making things simpler? As is sometimes recalled in philosophical discussions around the notion of simplicity (see, e.g., Norton, 2012a), Newton himself in the "Rules of Reasoning in Philosophy" of his *Principia* asserted that "[w]e are to admit no more causes of natural things than such as are both true and sufficient to explain their appearances" because "Nature does nothing in vain, and more is in vain when less will serve; for Nature is pleased with simplicity, and affects not the pomp of superfluous causes." But as Norton went on to show, simplicity is more clearly a reflection of our background knowledge than of Nature's modesty. Norton gave the following illustration: if we see, on a beach, a series of seagull steps in a line, the simplest assumption would be that these marks were left not by a group of birds each leaving one step's mark, but rather by a single bird walking a straight line; however, if we see a large number of the same steps not organized in a straight line, the simplest assumption would be that a flock of seagulls, and not just one disoriented bird, walked there. What makes an assumption the simplest in both cases seems be the number of causes—each bird leaves its own line of steps on the sand—rather than the number of birds.

There is no systematic way to identify what a theory should keep simple: the number of causes, of entities, of parameters? In the cosmology literature (Martin et al., 2011; Trotta, 2012) as well as in philosophy (Forster and Sober, 1994; Forster, 2000), but also

[6] One can find equivalent maxims in the cosmology literature. For instance, "the simplest theory compatible with the available evidence ought to be preferred." (Trotta, 2008, 130)

in the statistics literature itself (Spiegelhalter et al., 2002; Claeskens and Hjort, 2008, 2), information criteria are presented as a means to seek parsimony, i.e., a balance between complexity in terms of number of free parameters and fit or predictive accuracy of a model. Information criteria such as the BIC and the AIC result from different approaches, but they have similar effects, namely, for a given level of fit-to-data or predictive accuracy, they reward models with fewer parameters. Thus these criteria seem to weigh each parameter against its informativeness and penalize those parameters that only marginally increase fit. As we saw, that is why they are seen as a protection against overfitting, and why simplicity in terms of the number of parameters is assumed to play a role in model selection.

But we can see that, in model selection based on such information criteria, simplicity is not a relevant guiding criterion, even though the number of free parameter sometimes explicitly appears in the formulation with which these criteria are used. By having a closer look at the BIC and the AIC, we can see that the purpose and formulation of these information criteria have nothing to do with simplicity or balance between complexity and fit. We can then also see that more recent information criteria such as the Deviance Information Criterion (DIC) or the measure of Bayesian complexity (its analogue in the cosmology literature), which *explicitly* purport to measure a balance between complexity and fit don't necessarily depend on the number of parameters.

The Akaike Information Criterion is one of the most used information criteria, and one of the first to have been proposed (Akaike, 1973). The AIC of a given model is usually given in the following form:

$$-2\ln \mathcal{L}_{max} + 2k, \tag{5.2}$$

where $\mathcal{L}_{max}$ is the maximum likelihood achievable by the model and k the number of its free parameters. According to this criterion, the lower the score, the better the model.

For a given degree of fit, the AIC rewards models with fewer parameters, and vice versa. And because it also depends on the fit-to-data a model can achieve, it can been seen as balancing fit and complexity; this is why it is said to act as an Ockham's razor that limits the risk of overfitting. However, following (Norton, 2012b), we can see that the AIC is not defined in terms of simplicity, and that its referring to a model's dimensionality doesn't result from considerations about Ockham's razor or simplicity.

The AIC was first motivated by the Kullback-Leibler ("K-L") information (or divergence), defined as follows (see, e.g., Konishi and Kitagawa, 2008, §§ 3.1–3.4):

$$I(g; f) = \int \log\left(\frac{g(x)}{f(x)}\right) g(x)\,\mathrm{d}x, \tag{5.3}$$

where x is an observation (it can be a set of obervations) g is the true distribution for x and f a model. That model f, characterized by a $k-$dimensional parameter vector $\boldsymbol{\theta}$, is generally not identical to the true distribution. The K-L information measures how close the model f is to the true distribution g: its value is 0 if, and only if, the model f *is* the true distribution.

In practice, we don't know what the true distribution g is, and we want to find the model f that comes closest to it. To do so, we seek to maximize the following term of eq. (5.3):

$$\int g(x) \log f(x)\,\mathrm{d}x, \tag{5.4}$$

called the expected log-likelihood (since the other term of the integral, $\int g(x) \log g(x)\,\mathrm{d}x$, is fixed). But this term depends on the unknown distribution g. However, if we make the assumption that the data we have can provide us with a good estimate of the true distribution, then they can be used to determine the K-L information. To do so, we can replace the true distribution g with an empirical distribution $\hat{g}$ based on a set of N independent and identically distributed (IID) observations $\boldsymbol{x}_N$, where each of the N observations is weighted equally. This substitution is made under the assumption that this empirical distribution converges toward the true distribution when N is large. Now, in order to estimate the parameter vector $\hat{\boldsymbol{\theta}}$ of $\hat{g}$, based on the observations $\boldsymbol{x}_N$ and such that it maximizes the log-likelihood $\ell(\boldsymbol{\theta})$ of $f(\cdot|\boldsymbol{\theta})$ over all possible data, we can use analytic methods to obtain approximate solutions, called maximum likelihood method.[7] However, we introduce a bias by using the same data to estimate both the expected likelihood (eq. (5.4)) and the log-likelihood, $\ell(\boldsymbol{\theta})$, of our statistical model. Akaike found that, as $N \to \infty$, this bias comes asymptotically close to the model's number of free parameters.

Therefore, the AIC gives an evaluation of the badness (the bias) of a model whose parameters have been determined using the maximum likelihood method.[8] Improving a model's AIC score is equivalent to optimizing that model's log-likelihood over all possible data sets, and not just the data it was initially tuned for. But this implies that the true distribution $g(x)$ exists and is contained in the family of parametric models $\{f(x|\boldsymbol{\theta}); \boldsymbol{\theta} \in \Theta\}$ (where Θ is the entire parameter space) that we have initially specified. It also assumes that the hypothetical true distribution g is what generates the data we have access to.

Now, assuming that the unique, true distribution exists and is contained in the family of models under consideration, the AIC will automatically favor models of lower dimensionality. As Norton (2012b, 9–11) argued, if we expect the AIC to help us find a *unique* distribution, then it is appropriate for this information criterion to reward parsimony. Indeed, for some definite data, the same level of accuracy will be reached, within a given family of models, by a greater number of models of a given dimension than by models of lower dimension. Consider for instance Fig. 5.1, and assume that the data points reported in this figure are all the possible data points, without error, in a given context for which we are trying to find a true distribution. The 6th-order polynomial curve hits

[7] For instance, if $\ell(\boldsymbol{\theta})$ is continuously differentiable, the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ is a solution of $\dfrac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}$ (see Konishi and Kitagawa, 2008, § 3.3.2).

[8] The reason behind the addition of the -2 factor is debated (see, e.g., Rao et al., 2008, 352).

all these data points, and no curve of lower dimension does so. However, many curves of higher dimension could hit all the same data points, and an even larger number of curves of even larger dimension could do so. In this case, among all the distributions that hit all those points, the AIC will reward the model (i.e., the family of nth-order polynomials) with a unique solution, which is necessarily that with fewer parameters.

However, I claim—echoing Norton's remarks—that the fact that the AIC rewards models with fewer parameters is a byproduct of the analysis from which the AIC stems, rather than its explicit goal. The AIC, therefore, doesn't stem from considerations about parsimony, simplicity, or Aristotelian metaphysics. At least, it doesn't explicitly aim at achieving simplicity in terms of number of parameters as much as it aims at finding a unique, true solution.

We can see that the Bayesian Information Criterion, developed by Schwarz (1978) a few years after the AIC, doesn't depend on a model's dimension either. The BIC score for a given model is usually given in the following form:

$$-2\ln\mathcal{L}_{max} + k\ln N, \tag{5.5}$$

where $\mathcal{L}_{max}$ is the maximal likelihood achievable by the model (i.e., the Bayesian evidence of eq. (5.1) limited to that model), k the number of parameters in the model, and N the number of data points used to determine the fit. This assumes that the data points used are IID. The best model according to this criterion is the one with the lowest BIC score. In effect, like with the AIC, for a given level of fit-to-data, this model selection criterion rewards models that have fewer parameters. The BIC score depends on the sample size (i.e., N in eq. (5.5)), and in particular this has the effect of placing a higher standard for adding a parameter when the sample is higher, which can be seen as an appealing feature.[9]

[9]One might argue that the outcome of model comparison shouldn't depend on our sample size, but this effect vanishes for large sample sizes. Applications of the BIC must therefore be made with caution when the sample size is small. This assumption can be very problematic in cosmology in cases where we

However, balancing fit and complexity is *not* what the BIC *aims* at. In fact, the BIC isn't even defined in terms of a model's dimensionality. Indeed, its usual formulation, given above in eq. (5.5), is proportional to the approximation of the model likelihood (related to the Bayesian evidence given earlier in eq. (5.1)), for N observations. We can find the exact definition of the BIC in, e.g., (Konishi and Kitagawa, 2008, § 9.1.1). Assume that a model $\mathcal{M}_i$ is characterized by a parametric distribution $p_i(x|\boldsymbol{\theta}_i)$ and a prior distribution $p_i(\boldsymbol{\theta}_i)$ for a k_i-dimensional parameter vector $\boldsymbol{\theta}_i$. The likelihood of the model $\mathcal{M}_i$ for a set of N observations $\boldsymbol{x_N}$ ($\boldsymbol{x_N} = \{x_1, \ldots, x_N\}$ where the x_i are observations), is given by

$$p_i(\boldsymbol{x_N}) = \int p(\boldsymbol{x_N}|\boldsymbol{\theta}_i)\, p_i(\boldsymbol{\theta}_i)\, \mathrm{d}\boldsymbol{\theta}_i. \tag{5.6}$$

This assumes that the N observations are independent and identically distributed.

If there are r candidate models, Bayes theorem gives us the posterior probability of $\mathcal{M}_i$ ($i \in \{1, \ldots, r\}$):

$$p(\mathcal{M}_i|\boldsymbol{x}_N) = \frac{p_i(\boldsymbol{x}_N)\, p(\mathcal{M}_i)}{\sum\limits_{j=1}^{r} p_j(\boldsymbol{x}_N)\, p(\mathcal{M}_j)}. \tag{5.7}$$

If we assume that all the r models under consideration are equally likely (i.e., that their prior probability is identical), we can determine the posterior probability of $\mathcal{M}_i$ given N observations. It follows from eq. (5.7) that this is entirely determined by the model likelihood given in eq. (5.6). The BIC is defined as the logarithm of the model likelihood multiplied by -2:[10]

$$-2\ln p_i(\boldsymbol{x}_N) = -2\ln\left(\int p(\boldsymbol{x_N}|\boldsymbol{\theta}_i)\, p_i(\boldsymbol{\theta}_i)\, \mathrm{d}\boldsymbol{\theta}_i\right). \tag{5.8}$$

The lowest BIC score corresponds to the greatest model likelihood given N observations,

only have few (or sometimes only one!) observable events. We will come back to this in § 5.3.

[10]This formulation, and in particular the addition of the -2 factor, allows us to obtain a criterion close to the Akaike criterion, developed before the BIC.

which corresponds to the greatest model posterior given N observations and assuming that all the candidate models' priors are identical. The usual formulation for the BIC given in eq. (5.5), which depends directly on the model's dimensionality, results from the Laplace approximation of integrals (see Konishi and Kitagawa, 2008, § 9.1.2–9.1.3).

The BIC has interesting features. Among its advantages, it doesn't require us to integrate over the entire parameter space (as the Bayesian evidence requires), but only over a model's parameter space. In particular, for a given degree of fit, a model with fewer parameters will be rewarded, and vice versa. But this is not a *defining* feature of the BIC.

The AIC and BIC are the most widely used in model selection. We will see in § 5.3, however, that they are not well suited to most contexts in which we could carry out cosmological inferences. Ideally, cosmologists would prefer to work directly with the Bayesian evidence (see eq. (5.1)). But as we saw, the Bayesian evidence will almost always favor models that overfit the data. For that reason, they have suggested to add to the Bayesian evidence an "Ockham's razor penalty" (or "Bayesian complexity" score), so as to seek a balance between fit-to-data and complexity.[11]

For a model characterized by a $k-$dimensional parameter vector $\boldsymbol{\theta}$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, this information criterion makes use of the Kullback-Leibler distance, D_{KL}, between a parameter vector's prior, $p(\boldsymbol{\theta})$, and its posterior given the data, $p(\boldsymbol{\theta}|d)$:

$$D_{KL} = \int\limits_{\Theta} p(\boldsymbol{\theta}|d) \log \left(\frac{p(\boldsymbol{\theta}|d)}{p(\boldsymbol{\theta})} \right) \mathrm{d}\boldsymbol{\theta}. \tag{5.9}$$

[11] See, e.g., (Kunz et al., 2006; Trotta, 2008, 2012; Hobson et al., 2010, § 4–7). This model selection, the Bayesian complexity, is adapted from the Deviance Information Criterion (DIC): the explicit goal of this information criterion, given in the title of (Spiegelhalter et al., 2002) couldn't be clearer as to what its creators intended it to do. For a succinct but precise presentation of the DIC, see (Konishi and Kitagawa, 2008, § 9.5). See (Kunz et al., 2006; Hobson et al., 2010, § 4.3.2) for its application in cosmology.

The Bayesian complexity is defined as follows:

$$C = -2\left(D_{KL} - \widehat{D_{KL}}\right), \tag{5.10}$$

where $\widehat{D_{KL}}$ is the maximum information gain we can expect under the model (akin to the estimator for the AIC).[12] In other words, this information criterion measures how much parameter space is "wasted" in our model, and how well our data constrain our parameters. The closer to the truth a model, the smaller its parameter space, and the smaller its complexity score. As explained in (Hobson et al., 2010, 86), usually, in eq. (5.9), $D_{KL} \approx k \log\left(\frac{\text{signal}}{\text{noise}}\right)$, where k is the number of free parameters in our model, and where the "signal" and "noise" refer, respectively, to the relevant and wasted volumes in our parameter space.

Now, if we have two models $\mathcal{M}_1, \mathcal{M}_2$, with respective Bayesian evidences $p(D|\mathcal{M}_1)$ and $p(D|\mathcal{M}_2)$ and Bayesian complexities C_1 and C_2, Kunz et al. suggested to compare them as follows:

- $p(D|\mathcal{M}_2) >> p(D|\mathcal{M}_1)$: model $\mathcal{M}_2$ is clearly favored over model $\mathcal{M}_1$ and the increased number of parameters is justified by the data.
- $p(D|\mathcal{M}_2) \approx p(D|\mathcal{M}_1)$ and $C_2 > C_1$: the quality of the data is sufficient to measure the additional parameters of the more complicated model, but they do not improve its likelihood by much. We should prefer model $\mathcal{M}_2$, with less [*sic*] parameters.
- $p(D|\mathcal{M}_2) \approx p(D|\mathcal{M}_1)$ and $C_2 \approx C_1$: both models have a comparable likelihood and the effective number of parameters is about the same. In this case the data is not good enough to measure the additional parameters of the more complicated model and we cannot draw any conclusions as to whether the additional complexity is warranted. (Kunz et al., 2006,

[12] This is usually determined by using Markov chain Monte Carlo (MCMC) methods.

4)

Thus we see how the use of the Bayesian evidence and Bayesian complexity gives a formal characterization to the intuitive notion of "balancing fit and complexity." However, even if in some circumstances Bayesian complexity can be expressed in terms of the number of parameters, it is in fact defined by the volume of a model's parameter space, which depends on how the range of each parameter is defined. Now assume that we compare two models: $\mathcal{M}_1$, characterized by an $m-$dimensional parameter vector $\boldsymbol{\theta_1}$, and $\mathcal{M}_2$, characterized by an $m+1-$dimensional parameter vector $\boldsymbol{\theta_2}$. These models are of respective complexities C_1 and C_2. Assume further that $\mathcal{M}_1$ and $\mathcal{M}_2$ are identical, except that $\boldsymbol{\theta_1}$ includes a parameter α defined on a very broad range, and that, instead, $\boldsymbol{\theta_2}$ includes two parameters β, γ defined on very narrow ranges. We may then have $C_1 > C_2$ even though $\mathcal{M}_1$ has fewer parameters than $\mathcal{M}_2$.

Strictly speaking, then, simplicity—or a balance between simplicity and fit—is *not* what information criteria such as the Akaike information criterion or the Bayesian information criterion are after. And as we saw, strictly speaking, *none* of these information criteria are after simplicity defined in terms of number of parameters. This is even true of the Bayesian complexity criterion that cosmologists have introduced with the explicit intention of acting as an "Ockham's razor penalty" on the Bayesian evidence. Therefore, whether or not the use of such statistical tools is appropriate in general, and in cosmology in particular, does not depend on a sense of simplicity.[13]

There is a sense in which simplicity matters in model selection: we have to assume that *there is a true model*, and that there is but one true model—in other words, we have to assume that nature is not trying to trick us. In model selection and, more generally, in science, we strive to achieve parsimony in the number of models we should select. But that has nothing to do with parsimony with respect to the number of free parameters.

[13] One might argue that it is possible to find a notion of simplicity that captures what some or all of the information criteria seek. My claim concerns simplicity as it is usually addressed in statistics and cosmology (namely, simplicity in terms of number of parameters).

5.3 Limitations of Model Selection Criteria in Cosmology

The model selection criteria we saw above in § 5.2.2 can help us compare how well models fit a given body of data. As argued by Elliott Sober (2008, § 1.7), AIC and BIC are best used for different tasks: BIC favors models that give us the best likelihood (on average) for the data we have, whereas AIC favors models with the best predictive accuracy—that is, their ability to predict new data. In principle, in any situation where we would carry out model comparison with the help of the information criteria we saw previously, we could only compare the relative performance of models' fit or predictive accuracy and bias, but we couldn't assert that we found the true model based only on these methods. There may be relative advantages to the BIC or the AIC, but neither can tell us which one is true.[14] Only the AIC, however, can in principle tell us that, if the true model is among the family of models under consideration, we have come as close to it as possible.

"Cosmologists, however, are probably not yet willing to concede that they might be looking for something other than absolute truth specified by a finite number of parameters." (Liddle, 2007, 76) But predictive accuracy, assessed by the AIC, cannot suffice to characterize "closeness to the truth," contrary to claims one can find, e.g., in (Forster and Sober, 1994, 10). Indeed, *false models can be predictively accurate*—think of the Ptolemaic model evoked in the previous section for instance.[15] Therefore, in cosmology where the number of relevant parameters is not settled, neither maximal fit nor best predictive accuracy, nor, as we saw, balance between complexity and fit, will suffice as model selection criteria.

[14] To be sure, this is not the aim of these methods. It is beyond the scope of this paper to characterize all the relative advantages and limitations of these information criteria. See (Liddle, 2007) for a discussion about the relative merits of the BIC and AIC. The former generally avoids overfitting better than the AIC, but it can also unfairly disfavor more complex models, and the risk of overfitting can be replaced by that of underfitting; see also (Weakliem, 1999) for a discussion of this problem from a social science perspective.

[15] In this paper, I have not always carefully distinguished between "predictive accuracy" and "fit", and considered both as a type of fit-to-data, whether old or future.

What we can expect from such statistical methods is that they help us to select which of our candidate models has the best average likelihood or the best predictive accuracy with the least bias, not which are the candidate models. However, we can see that, even if we assume that among the models we compare one of them *is* the true underlying distribution, then, there are in cosmology limitations to their meaningfulness in principle. The two main obstacles for the application of these model selection methods in cosmology are 1) the lack of well-motivated prior distributions for some cosmological parameters,[16] and 2) the limited accessible sample of cosmological events—i.e., the uniqueness of the universe.

Model comparison based on the Bayesian evidence (such as the "balance-between-fit-and-complexity" approach proposed by Kunz et al.) assumes that we can assign prior probabilities to the parameters at play in our models. However, our theoretical background leaves some cosmological parameters undetermined; Efstathiou (2008) for instance showed how this is the case for dark matter. As a consequence, the Bayesian evidence can be highly dependent on some arbitrary choices regarding how we define the parameters' prior distributions. For instance, even if we give a relatively much higher prior probability to a small range of values around the center of a normal distribution, the outer tails of such a probability distribution can contain much of the volume of the parameter space. In model selection, this effect becomes exponentially problematic as the number of parameters increases, up to the point where almost all of the volume can be located the extreme ranges of the parameter space.[17] Therefore, comparing models whose dimensionality differ can be very risky. Model comparison, whether based on Bayesian evidence or other criteria, doesn't replace the parameter inference step, and, when our background knowledge doesn't give us much constraint for some parameters, arbitrary choice of priors for them can therefore greatly affect a model's Bayesian evidence.

[16] This was argued in (Linder and Miquel, 2008) and, more cogently, in (Efstathiou, 2008).

[17] Up to 99.8% for a 10-dimensional space. See, e.g., http://astrostatistics.psu.edu/su05/bayes_freq2.pdf, cited in (Linder and Miquel, 2008).

Consider for instance the case of the comparison of a cosmological model with a flat universe (i.e., in which the curvature Ω_κ is null) with a curved model ($\Omega_\kappa \neq 0$). To do so, we would determine the Bayes factor of two models:

- a model $\mathcal{M}_1$ such that $p(\Omega_\kappa|\mathcal{M}_1)$ (the prior distribution for Ω_κ in $\mathcal{M}_1$) is a Gaussian distribution of mean $\mu = 0$ with a given variance τ^2, and

- a model $\mathcal{M}_2$ such that $p(\Omega_\kappa|\mathcal{M}_2)$ is a Gaussian distribution of mean $\mu \neq 0$ with a given variance τ^2.

A comparison based on the ratio of the Bayesian evidence will require us to determine, for each model, the data likelihood given Ω_κ as well as the prior distribution for Ω_κ in that model (where $\Omega_\kappa = \theta$ in eq. (5.1)). Now, as illustrated in (van Dyk, 2012, 144),[18] the Bayesian evidence is highly dependent on the prior distribution $p(\Omega_\kappa)$; in particular, depending on our choice of *variance* τ^2 for that parameter, the Bayes factor can strongly favor either model!

We saw earlier that cosmologists sometimes appeal to the "Astronomer's prior," a uniform prior distribution for $-1 \leq \Omega_\kappa \leq 1$, since it is consistent with observable properties of the universe. But based on theoretical considerations from inflationary cosmology, the "Curvature scale prior," a uniform prior distribution for $-5 \leq \log|\Omega_\kappa| \leq 0$, would have produced different results. Our results are therefore strongly dependent on an arbitrary choice of priors and on the lack of uninformative priors.[19] This dependence on a choice of prior undermines the appeal of likelihoodist methods, which seeks to avoid the need for prior probability distribution for the models being compared.

If the Bayesian evidence depends on such arbitrary choices, then so will the Bayes factor. Moreover, not only will comparing the Bayesian evidence's fit with a complexity factor dependent on these choices, but also the complexity factor itself is not invariant

[18] which constitutes a response to (Trotta, 2012).

[19] As van Dyk (2012, 144) recalled, the use of improper prior distributions would result in improper prior predictive distributions and undefined Bayes factor.

under reparametrization.[20] Likewise, even though the BIC only requires us to estimate the maximum likelihood achievable *within a given model*, and not throughout the parameter space, its results may also be strongly dependent on an arbitrary choice of prior distribution—and especially the parameter range—for a parameter within that model (see Weakliem, 1999; Liddle, 2007).

The AIC, on the other hand, isn't affected by how we set *a priori* the parameters' range. But we can see that in cosmology its application and that of the BIC face limitations that are inherent to the fact that in cosmology we only have access to *one realization* of a cosmological model—i.e., one universe. Here, I don't want to reiterate claims that can be found in, e.g., (Ellis, 2007) according to which the uniqueness of the universe may preclude the possibility of cosmology as a science; the problem I want to underline concerns the use of *statistical* methods for analyzing a unique data set about a *unique* object.

Indeed, statistical methods such as those we have considered so far are usually applied to a set of multiple realizations of a given process or phenomenon. But there are in cosmology limitations that our sample can't overcome:

1. There is an inherent variance in our data due to the fact that we can only perform observations from one point of view. That's a sample variance due to our observation bias and technical limitations, for instance in techniques used to determine the power spectrum of the Cosmological Microwave Background radiation (CMB) with harmonics.[21] As a consequence, there is a limiting precision in how our data can be used as evidence for one model or another (e.g., anisotropies in the CMB can determine the amplitude of primordial mass fluctuations).

2. There is another kind of sample variance: if the physical process underlying the formation and evolution of the universe will do so with some level of randomness or

[20]See, e.g., (Spiegelhalter et al., 2014) for a discussion of this problem.
[21]See the large error bars at large angles below in Fig. 5.2.

variation, then we are limited by the fact that we can only access *one* realization of that process.

This first limitation hinders our capacity to access independent measurements to the same phenomenon. But in observations of the CMB, for instance, even if we assume no error in our measurement at all (the kind of error that independent measurements would reduce), our data underdetermines the power spectrum of the CMB at large angular scales.[22]

There are solutions to overcome the first kind of limitation at low angular scales, by seeking to observe the CMB from "other viewpoints", for instance through a cluster of galaxies (see Kamionkowski and Loeb, 1997).[23] But such solutions can't address the second kind of limitation, which produces an uncertainty about a variance in our data at both low and large angular scales.

These problems affect the validity of the results we could obtain with model selection information criteria in different ways. We saw for instance that, independently of our limited background knowledge, and independently of the problem of defining parameters and their range (in general or even only in a given model), a model's BIC score depends on our sample size (see eq. 5.5-5.6), especially if this sample size if small. Likewise, the AIC score will be affected by the sample size, and its definition in terms of the model's number of free parameters comes from an approximation for a large number of IID observations.

But the consequences of the limitation of our sample are even more problematic for the AIC. Indeed, the AIC assumes that the data is generated by one of the models under consideration. That is, it assumes that the AIC evaluates how biased a given model is from the true model, assuming that the true model can be found in the family of models we have specified (by identifying candidate parameters), and assuming that this true

[22] This problem is often referred to in the literature as "cosmic variance", even though this term can be used to refer to other of the sampling problems discussed here.

[23] In a nutshell, one can observe the polarization of CMB photons scattered by electron gas found in a cluster of galaxies; this polarization corresponds to properties of the CMB as seen in this cluster.

model is what generates the data from which we carry out this statistical investigation—that is, we assume that the data we have aren't itself error-laden. It is under this assumption that, as we saw with eq. (5.4), we can substitute the average of our data, $\hat{g}$, for the true distribution g. Thus the very idea behind the AIC assumes that have at our disposal a large sample of IID data sets that we take as a surrogate for the true distribution from which the K-L information measures the distance. If, however, there is only one data set from which we derive both this surrogate and our model, then the AIC can't do more than give us an estimate of goodness-of-fit.

Methods usually presented as allowing us to get around the cosmic variance, such as in (Kamionkowski and Loeb, 1997), allow us to obtain independent observations, but independent observations *of the same event*. The kind of IID observations statistical methods rely on are observations of different, independent *iterations*, or realizations, of the same physical process.[24] These methods were not designed—and their formulation would not be as well justified—for small samples, let alone a sample of one event as is the case with the CMB!

To be sure, there are contexts in which the anisotropies in the CMB can be each treated as IID events. For instance, if we want to use fluctuations in the CMB as evidence regarding how the early universe evolved at later stages, or if we are interested in studying whether these fluctuations are related to present-day large-scale astrophysical structures, then each of these anisotropies, at different angular levels, and each of these structures, can be considered as IID events. But if we want the CMB to inform us on the physics responsible for these anisotropies—i.e., truly cosmological properties—then together they constitute only one event.

Therefore, without being able to use Bayesian information criteria under adequate circumstances, and without even the hope of improving these circumstances, we place ourselves in a situation where we can't know, with these methods, if the model they

[24]The cosmology literature is sometimes explicit about this distinction, but, as was noted in a similar context, "this level of sloppiness is standard." (Starkman et al., 2012, 4)

would favor isn't overfitted to too small a sample, which would defeat the very purpose of using these information criteria in the first place.

This problem will arise, for instance, in the analysis of the CMB. The CMB is a highly isotropic radiation which, according to the big bang model, dates from the epoch of recombination in the early universe, when photons could first travel freely without interacting with matter. The angular power spectrum represents the small anisotropies in the CMB as a function of angular distance in the sky. Such observations of anisotropies in the CMB constitute a major if not the main empirical test of cosmological models and are taken to provide crucial constraints on cosmological parameters (see, e.g., Parkinson and Liddle, 2010; Martin et al., 2011). Fig. 5.2 shows the angular power spectrum obtained from WMAP data: the continuous line is the best-fitting model from inflationary cosmology; the large error bars at large angular (i.e., low-l) scales are due to cosmic variance (in the first sense given earlier). But even with this limitation in precision, significant anomalies exist at large angles; the angular power spectrum is significantly lower at large angles than predicted by theory.[25]

As George Ellis (2014, § 3.1 and p. 12) recalled, a lower power spectrum at those scales are expected to occur in a "small universe". Whether or not these anomalies are physically meaningful (i.e., are not artifacts) will affect what we can tell about the large-scale topology of our universe, and therefore our choice of cosmological models.

These discrepancies are usually considered to be "statistical flukes" that reveal the limit of using statistical methods in such analyses for a unique event (see, e.g., Copi et al., 2007, for a discussion). Therefore, if indeed these anomalies are due to observational artifacts, then trying to account for them at all cost would result in a model that overfits the data.

With a similarly small sample of events at other angular scales, we could as well disregard anomalies (for instance, at $l \approx 40$) as being statistical flukes as well. This,

[25] This discrepancy persists in more recent observations made by Planck (see Ijjas et al., 2013).

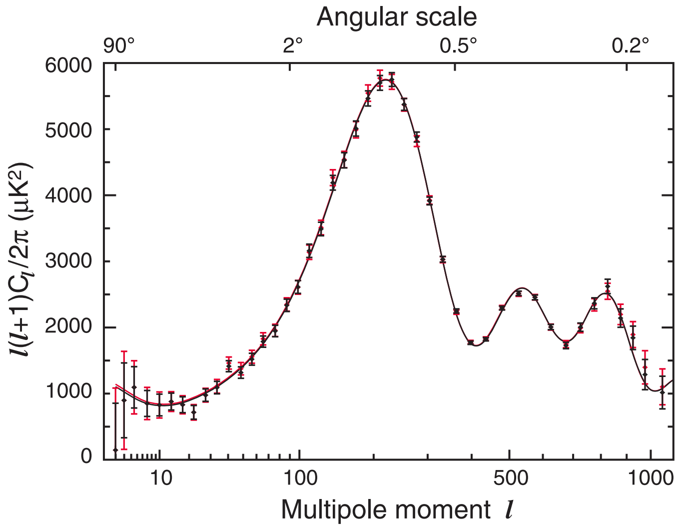

Figure 5.2: Two estimates (in red and in black) of the WMAP angular power spectrum of the Cosmic Microwave Background radiation. The continuous lines are the best-fitting models from inflationary cosmology for each estimate. Source: NASA/WMAP.

however, cannot be the only argument by which we discard problematic disagreements between data and theory, for it would be ad hoc, especially if these anomalies could be explained by equally plausible scenarios (see, e.g., Liu et al., 2013). However, the kind of model selection methods considered in earlier sections will *not* be able to tell us what data point in our small sample is a statistical fluke or not. These methods cannot choose models for us, nor can they tell us how to interpret our data; they can only help us compare how well these models fit the data and, *assuming* that one of these models is true, how close to the truth it is.

5.4 Measuring Consilience Rather Than Predictive Accuracy or Fit

We have seen that the difficulties to use Bayesian methods for model selection in cosmology are not easy to overcome. We saw in § 5.2.1 that likelihoodist methods that rely on the Bayesian evidence or the Bayes factor can be highly dependent on a choice of priors—

at least a choice of range—for our physical parameters and, if the number of parameters is not fixed a priori, they run the risk of overfitting the data. The BIC and the AIC, in different ways, aim at addressing both problems. First, the BIC is less sensitive than the Bayes factor to the parameter range problem, and the AIC avoids that problem entirely. Secondly, both criteria are usually introduced as solutions to avoid overfitting by seeking parsimony in terms of our models' number of free parameters. In § 5.2.2 we have seen, in fact, that the validity and applicability of these criteria to cosmology does not depend on a vaguely defined notion of simplicity or ontological parsimony. But we saw in § 5.3 that, regardless of the role of simplicity in model selection, there are limitations to the use and meaningfulness of these criteria in cosmology. In particular, the BIC and the AIC were designed to apply to large samples of IID observations. However, in many cases in cosmology, in addition to being mostly limited to observing from one particular standpoint, we can only hope to be able to observe a unique event. It is questionable, under these circumstances, to use criteria *defined under the assumption that they will apply to large samples*. Furthermore, it undermines the relevance of these criteria if we use them in situations in which we can't identify which of our data, if any, are statistical flukes, since this can greatly affect our conclusions.

We can see that Bayesian methods can play a role in model selection in cosmology and overcome some of these difficulties. However, I will suggest that it is not to assess fit-to-data, predictive accuracy, or balance between fit and complexity for which Bayesian methods will be most legitimate in cosmology. Even with limited data about a unique event, we can assess a model's ability to unify phenomena and physical processes that may otherwise seem unrelated. Newton's theory of gravitation, for instance, gave a common causal framework to both astronomical and terrestrial phenomena. Another example sometimes given in support of the view of induction as unification is Jean Perrin's case for atomism; by providing several experimental methods to calculate Avogadro's number from disparate domains—Brownian motion, the viscosity of gases, the color of the sky, black body spectrum, etc.—Perrin argued that this concordance of evidence could not be "be considered as the result of chance." (quoted in Psillos, 2011, 355)

Such cases of concordance and explanatory and causal unification illustrate William Whewell's view of "consilience" as a confirmation criterion:

> the evidence in favour of our induction is of a much higher and forcible character when it enables us to explain and determine cases of a *kind different* from those which were contemplated in the formation of our hypothesis. (...) That rules springing from remote and unconnected quarters should thus leap to the same point, can only arise from *that* being the point where truth resides.
>
> Accordingly the cases in which inductions from classes of facts altogether different have thus *jumped together*, belong only to the best established theories which the history of science contains. And as I shall have occasion to refer to this particular feature in their evidence, I will take the liberty of describing it by a particular phrase; and will term it the *Consilience of Inductions.* (Whewell, 1847, 65, cited in (Myrvold, 2003))

Induction for Whewell consists in several steps, namely the identification of the independent variables, the construction of models, and the determination of parameters. Consilience then occurs when the value of parameters determined from a kind of phenomena agrees with that obtained from a distinct kind of phenomena. In other words, consilience occurs when we multiple, independent bodies of data used to determine the values of parameters contained in our models. Conversely, in order to check that consilience occurs, we can use the determination of parameters from one body of data to place constraints on that other body of data. A model that suggests such a lawlike connection between disparate bodies of data therefore makes them informationally relevant to each other.[26]

[26] "Consilience" is stronger than mere "concordance" as it is used in the cosmology literature, to

Wayne Myrvold (2003) gave a Bayesian account of consilience. Among the ingredients of this account is the notion of *informational relevance* $I(P, Q)$ *of one proposition* P *to another proposition* Q, that is, the information that one proposition P yields about another proposition Q, formulated as follows:[27]

$$I(P, Q) = \log \frac{p(Q|P)}{p(Q)}. \tag{5.11}$$

This measure is defined such that independent evidence is additive. That is, if P_1 and P_2 are independent, then $I(P_1 \& P_2, Q) = I(P_1, Q) + I(P_2, Q)$. Moreover, it is symmetric with respect to its arguments P and Q.

The second ingredient of this account is a *measure of unification* $U(P_1, P_2; H)$ of two phenomena P_1, P_2 by a hypothesis H, that is, how much H makes P_1 yield information about P_2, defined as follows:

$$U(P_1, P_2; H) = I(P_1, P_2|H) - I(P_1, P_2). \tag{5.12}$$

Finally, we can define a measure of evidential support $I(H, P_1 \& P_2)$ of two phenomena P_1, P_2 to a hypothesis H on the basis of its unificatory power, as follows:

$$I(H, P_1 \& P_2) = I(H, P_1) + I(H, P_2) + U(P_1, P_2; H) \tag{5.13}$$

What this measure captures is the fact that the unificatory power of a hypothesis, by

designate a region of parameter space where ΛCDM models matches all the data we have. The latter only refers to empirical constraints, whereas consilience is about mutual informational relevance between two bodies of data otherwise unrelated.

[27] $I(P, Q)$ is assumed to be continuously definable—in terms of probability functions for Q and Q given P—so that small changes in probability yield small changes in information. As explained in (Myrvold, 2003, 409–410), it also follows a normalization convention such that, when our information about Q amounts to certainty that Q obtains, "[i]f Q is one of 2^n equiprobable, mutually exclusive and jointly exhaustive alternatives, then the information that Q obtains amounts to n bits of information, and, in general, information that Q obtains will count as $-\log_2 p(Q)$ bits of information." (notation edited for consistency) Assumed here and in the other elements of this account is the fact that these probability functions are defined with respect to our background information.

itself, contributes to its support by the evidence. This measure differs from the information criteria seen above in several ways. This measure of consilience allows us to assess how much a physical model, characterized by a set of physical relationships between some parameters, unifies phenomena. The BIC and criteria based on the Bayesian evidence, on the other hand, allow us to assess the goodness-of-fit of a model defined as a set of probability distributions for the values of some parameters, and the AIC the bias of such models and their predictive accuracy. All this measure of consilience requires—and all it assesses—is the ability of a hypothesis under consideration to make two sets of phenomena yield information about each other.

This alternative information criterion is interesting in cosmology because it doesn't suffer from some of the limitations of model selection criteria we saw earlier in § 5.3. To be sure, the information that a model yield in virtue of its unificatory power still depends on a possibly arbitrary choice of prior distributions for our parameters. But, unlike the AIC, this Bayesian measure of consilience doesn't require in principle that we have access to multiple, IID data sets. Indeed, the probability distributions at play in eq. (5.11, 5.13) need only reflect our credence for a given parameter or hypothesis, whatever the size of our sample. Because, on this account, models are assessed according to how well they connect independent bodies of data, we avoid the risk of overfitting our models to what might have been indistinguishable statistical flukes. Moreover, this measure of consilience allows us to favor a model over another based on empirical considerations, and not on vague, extra-empirical intuitions about "Nature's modesty" or maxims allegedly from medieval thinkers.

William Harper (2011) has given an exposition of how this account of empirical success fits Newton's methodology. An illustration of this concerns how Newton determined the exponent of the power law with observations of pendulums, the motion of planets and that of the apsides. In doing so he used independent observations to constrain the same parameter value from one context to another. Harper also showed how it continues to guide cosmology today: it is only when multiple, independent bodies of data were increasingly and more precisely agreeing with each other on the existence and the value of dark energy that this became an established fact to most cosmologists (see Harper, 2011, 394–396).[28]

Relatedly, we can assess the unificatory power of dark matter from different sources: from the direct or indirect detection of weakly interacting massive particles (WIMPs), from astrophysical sources or from particle colliders; from the statistical distributions of galaxies and their rotation curves; from distant supernovæ; from lensing effect, etc. Likewise, observations of fluctuations in the CMB can constrain models of WIMP (see Bauer et al., 2015).

The idea that consilience should be used as a criterion of confirmation may not be new or surprising to cosmologists, yet it can offer new perspectives on how to approach longstanding problems, such as confirming inflationary cosmology. A shift in confirmation criteria from simplicity to consilience would have the advantage, for instance, not to reward the versatility of inflationary cosmology—i.e., its ability to generate models that could accommodate nearly any initial condition in the early universe.[29] Chris Smeenk (2012), for instance, argued that considering inflationary cosmology as a a theory of structure formation, rather than as a solution to fine-tuning problems in the early universe, offers more detailed constraints on its parameters by relating the amplitude of the density perturbations of the inflaton field to various features of the CMB.[30]

With these few fragmentary remarks on confirming aspects of dark energy, dark matter, or inflationary cosmology, I am only hinting at why a measure of empirical unification and theoretical consilience constitutes a better approach to cosmological model selection

[28] This took place after the first measurements on the CMB from COBE, and more decisively those of WMAP, corroborated estimations of $\Omega_\Lambda \approx 0.7$ from a previous survey of supernovæand, indirectly, from estimates of the age of the universe by observing the oldest clusters of galaxies.

[29] One can find criticisms of the versatility of inflation in the cosmology literature, but they are usually phrased, vaguely, in terms of Popperian falsifiability, which is arguably a coarse confirmation criterion.

[30] The difficulties in trying to characterize inflation as a unification between between cosmology and particle theory, which would constitute an alternative attempt to characterize the potential unifying power of inflation, was discussed in (Zinkernagel, 2002).

than the statistical and Bayesian methods we saw previously. I here mainly want to make the case that we can use a quantitative measure of confirmation which, in cosmology, is better motivated and doesn't suffer from the same limitations as the model selection methods often used. Consilience and unification as a confirmation criterion is more demanding than predictive accuracy, and more refined than falsifiability.

5.5 Conclusions

We have seen that when the theoretical landscape is not well-defined or well-constrained, comparing cosmological models based on their respective Bayesian evidence is not a reliable method, as it is likely to favor models that overfit the data. To avoid this, we have seen attempts in the literature to appeal to simplicity, given in terms of our models' number of free parameters: cosmologists have proposed a measure of "Bayesian complexity" as a correction put on a model's Bayesian evidence so as to penalize that model's complexity. Alternatively, the cosmology literature sometimes appeals to other statistical and Bayesian information criteria such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC), often under the assumption that their goal is to find a balance between fit-to-data and complexity. In the literature, this appeal to simplicity is often justified by vague, apocryphal references to Ockham's razor and to ontological restraint.

A closer look at these model selection criteria has shown that, in spite of how they are usually formulated or presented, they do *not*, strictly speaking, aim at favoring models that, other things being equal, have fewer parameters. Rather, these criteria's formulations in terms of a model's number of free parameters only hold under certain conditions and approximations—assumptions or approximations with respect to the size of our sample or to our parameters' range. Therefore, as Norton (2012b) has shown for the AIC, simplicity with respect to the number of free parameter is more a common byproduct than a goal or a consistent result of the analysis carried out with any of these model selection criteria.

Although, in practice, these model selection criteria tend to sanction the introduction of new parameters only if it results in a considerable gain in information about the data, it is objectionable to say that it is out of a sense of parsimony or ontological modesty that we use these criteria. Not only is there no straightforward or objective sense in which to characterize simplicity, but also is it doubtful that Nature obeys any imperative to be modest or simple—at least, there is no compelling reason why it should be simple with respect to its number of free parameters.

Now, regardless of metaphysical considerations about these model selection criteria, we have seen that the use of these criteria is especially problematic in cosmology. Indeed, conclusions based on the BIC score or the Bayesian evidence can be highly dependent on arbitrary choices of priors or on how parameter ranges are defined. Moreover, accuracy in using the AIC (and, to a lesser extent, the BIC) assumes that we have access to a large sample of independent and identically distributed data. But this requirement cannot be met if these model selection techniques are used to characterize a unique object (such as the universe) or a unique event (such as the CMB).

I have suggested, however, that another Bayesian measure could help us assess and compare parameters and models even when our data sample is very limited. The measure of consilience or unification proposed by Myrvold (2003)—that is, a measure of a model's ability to unify different sorts of phenomena—does not suffer from the some of the limitations that undermine the use of other model selection methods in cosmology. By measuring a model's ability to make two otherwise unrelated sets of phenomena yield information about each other by giving a functional relationship between them, characterized by some parameters, this model selection method does not merely select the "best" set of parametric values according to a sense of balance between fit and simplicity. The degree of confirmation from the unificatory power of a model it assesses, although it depends on assumptions about our prior expectations, allows us to ground model comparison in empirical claims rather than scholastic metaphysics.

Bibliography

Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. In *International Symposium on Information Theory*, pages 267–281.

Ariew, R. (1977). Did Ockham Use His Razor? *Franciscan Studies*, 37(1):5–17.

Bauer, D., Buckley, J., Cahill-Rowley, M., Cottad, R., Drlica-Wagner, A., Feng, J. L., Funk, S., Hewett, J., Hooper, D., Ismail, A., Kaplinghat, M., Kusenko, A., Matchev, K., McKinsey, D., Rizzo, T., Shepherd, W., Tim M.P. Tait, Wijangco, A. M., and Wood, M. (2015). Dark Matter in the Coming Decade: Complementary Paths to Discovery and Beyond. *Physics of the Dark Universe*, 7-8:16–23.

Claeskens, G. and Hjort, N. L. (2008). *Model Selection and Model Averaging*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, UK.

Copi, C. J., Huterer, D., Schwarz, D. J., and Starkman, G. D. (2007). Uncorrelated Universe: Statistical Anisotropy and the Vanishing Angular Correlation Function in WMAP Years 1-3. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, 75(2):1–17.

Efstathiou, G. (2008). Limitations of Bayesian Evidence applied to cosmology. *Monthly Notices of the Royal Astronomical Society*, 388:1314–1320.

Ellis, G. F. R. (2007). Issues in the Philosophy of Cosmology. In Butterfield, J. and Earman, J., editors, *Handbook of the Philosophy of Science, Philosophy of Physics, Part B*, pages 1183–1283. Elsevier, Amsterdam.

Ellis, G. F. R. (2014). On the Philosophy of Cosmology. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 46:5–23.

Forster, M. R. (2000). Key Concepts in Model Selection: Performance and Generalizability. *Journal of Mathematical Psychology*, 44(1):205–231.

Forster, M. R. and Sober, E. (1994). How to Tell when Simpler, More Unified, or Less Ad Hoc Theories will Provide More Accurate Predictions. *The British Journal for the Philosophy of Science*, 45(1):1–35.

Harper, W. L. (2011). *Isaac Newton's Scientific Method: Turning Data into Evidence about Gravity and Cosmology*. Oxford University Press, New York.

Hitchcock, C. and Sober, E. (2004). Prediction Versus Accommodation and the Risk of Overfitting. *The British Journal for the Philosophy of Science*, 55:1–34.

Hobson, M. P., Liddle, A. R., Mukherjee, P., and Parkinson, D., editors (2010). *Bayesian Methods in Cosmology*. Cambridge University Press, Cambridge.

Ijjas, A., Steinhardt, P. J., and Loeb, A. (2013). Inflationary Paradigm in Trouble After Planck2013. *Physics Letters B*, 723(4-5):261–266.

Kamionkowski, M. and Loeb, A. (1997). Getting Around Cosmic Variance. *Physical Review D*, 58(8):4511–4513.

Konishi, S. and Kitagawa, G. (2008). *Information Criteria and Statistical Modeling*. Springer, New York.

Kunz, M., Trotta, R., and Parkinson, D. R. (2006). Measuring the Effective Complexity of Cosmological Models. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, 74(2):1–9.

Liddle, A. R. (2004). How Many Cosmological Parameters? *Monthly Notices of the Royal Astronomical Society*, 351(3).

Liddle, A. R. (2007). Information Criteria for Astrophysical Model Selection. *Monthly Notices of the Royal Astronomical Society*, 377(1):L74–L78.

Liddle, A. R. (2009). Statistical Methods for Cosmological Parameter Selection and Estimation. *Annual Review of Nuclear and Particle Science*, 59:95–114.

Linder, E. and Miquel, R. (2008). Cosmological Model Selection: Statistics and Physics. *International Journal of Modern Physics D*, 17(12):2315–2324.

Liu, Z. G., Guo, Z. K., and Piao, Y. S. (2013). Obtaining the CMB Anomalies with a Bounce from the Contracting Phase to Inflation. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, 88(6):1–9.

Longair, M. L. (1993). Modern Cosmology: A Critical Assessment. *Quarterly Journal of the Royal Astronomical Society*, 34(2):157–199.

Martin, J., Ringeval, C., and Trotta, R. (2011). Hunting Down the Best Model of Inflation with Bayesian Evidence. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, 83(6):1–11.

Myrvold, W. C. (2003). A Bayesian Account of the Virtue of Unification. *Philosophy of Science*, 70(2):399–423.

Norton, J. D. (2012a). Simplicity as a Surrogate. In *The Material Theory of Induction (forthcoming)*, chapter 6, pages 1–46.

Norton, J. D. (2012b). Simplicity in Model Selection. In *The Material Theory of Induction (forthcoming)*, chapter 7.

Parkinson, D. R. and Liddle, A. R. (2010). Application of Bayesian Model Averaging to Measurements of the Primordial Power Spectrum. *Physical Review D*, 103533(September):1–8.

Psillos, S. (2011). Moving Molecules Above the Scientific Horizon: On Perrin's Case for Realism. *Journal for General Philosophy of Science*, 42(2):339–363.

Rao, C., Toutenburg, H., Shalabh, H. C., and Schomaker, M. (2008). *Linear Models and Generalizations: Least Squares and Alternatives.* Springer, 3rd edition.

Schiffrin, J. S. and Wald, R. M. (2012). Measure and Probability in Cosmology. *Physical Review D*, 86(2):023521.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464.

Smeenk, C. (2012). Inflation and the Origins of Structure (unpublished manuscript).

Sober, E. (2008). *Evidence and Evolution: The Logic Behind the Science.* Cambridge University Press.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian Measures of Model Complexity anf Fit. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 64(4):583–639.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van der Linde, A. (2014). The Deviance Information Criterion: 12 Years On. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 76(3):485–493.

Starkman, G. D., Copi, C. J., Huterer, D., and Schwarz, D. J. (2012). The Oddly Quiet Universe: How the CMB Challenges Cosmology's Standard Model. *Romanian Reports of Physics*, 57(5-6):979–991.

Szydlowski, M., Krawiec, A., Kurek, A., and Kamionka, M. (2015). AIC, BIC, Bayesian evidence against the interacting dark energy model. *The European Physical Journal C*, 75(5):1–11.

Thorburn, W. (1918). The Myth of Occam's Razor. *Mind*, 27(170):345–353.

Trotta, R. (2008). Bayes in the Sky: Bayesian Inference and Model Selection in Cosmology. *Contemporary Physics*, 49(2):71–104.

Trotta, R. (2012). Cosmological Bayesian Model Selection: Recent Advances and Open Challenges. In Feigelson, E. D. and Babu, G. J., editors, *Statistical Challenges in Modern Astronomy V*, volume 902 of *Lecture Notes in Statistics*, pages 127–140. Springer New York, New York, NY.

van Dyk, D. A. (2012). Commentary: Cosmological Bayesian Model Selection. In Feigelson, E. D. and Babu, G. J., editors, *Statistical Challenges in Modern Astronomy V*, Lecture Notes in Statistics, chapter 12, pages 141–146. Springer, New York, NY.

Weakliem, D. L. (1999). A Critique of the Bayesian Information Criterion (BIC) for Model Selection. *Sociological Methods and Research*, 27(3):359–397.

Whewell, W. (1847). *The Philosophy of the Inductive Sciences, Founded upon Their History.* J.W. Parker, London.

Zinkernagel, H. (2002). Cosmology, Particles, and the Unity of Science. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 33(3):493–516.

Chapter 6

Concluding Remarks

I will end this dissertation with a few concluding remarks, so as to emphasize its contributions for philosophy and cosmology.

Three of the papers in this work assessed anthropic arguments and predictions *à la* Weinberg, and particularly its two central premises: a representation of ignorance or indifference, and the claim that we can conditionalize a probability distribution for a physical parameter on the number of observers it allows for.

The first two papers, Chapters 2 and 3, focused on the first of these premises, i.e., the problem of reasoning from ignorance or indifference, and argued that the ability to draw conclusions in such a situation is a flaw rather than a strength of the use of probabilities. In Chapter 3, I have shown that, under these circumstances, such flawed arguments and conclusions (of which the Doomsday argument is another example) are necessary consequences of any Bayesian approach that represents credal states by single credence functions, unless we adopt ad hoc rules as the ones we can find in (Bostrom, 2002a). But I have also shown that these conclusions can be avoided if we allow our credences to be imprecise.

What is at stake with such arguments is the adequacy of probabilities as an inductive framework, recently summarized as follows:

> John Norton has discussed the limits of probability theory as a logic of induction, using an example which, he claims, admits no reasonable probabilistic

> attitude (Norton, 2007, 2008a,b). One might hope that [the imprecise probability model] offers an inductive logic along the lines Norton sketches. Norton himself has expressed scepticism on this line (Norton, 2007). (Bradley, 2015, § 2.6)

In particular, Norton had put forth compelling criteria any logic of induction should meet when representing indifference or ignorance. In response to Norton's skepticism, I have shown, in Chapter 2, that imprecise probabilities can indeed meet these criteria. Thus I hope to have convincingly argued not only that the imprecise credence model constitutes a plausible framework for an inductive logic, but also that it achieves what no other representation of credences by a unique probability distribution can do.

The third paper examined the second premise in Weinberg's anthropic argument, i.e., the claim that we ought to expect to find ourselves in a physical world more favorable to the advent of life. This claim is sometimes—even in the cosmology literature—formulated as one about self-locating uncertainty, to which solutions to the Sleeping Beauty problem are relevant. I have made precise what a defense of anthropic reasoning based on the Sleeping Beauty problem or, more generally, about self-locating uncertainty, can be. But at the same time I have shown that the two kinds of arguments are, in fact, not similar.

This examination of the Sleeping Beauty problem led me to argue against the "thirder" position of Elga et al., and thus to dispute that this problem can be used to illustrate that, in science in general and cosmology in particular, "we need a methodology for evidence with a *de se* component." (Bostrom, 2002b, 621) More precisely, I presented three arguments against the "thirder" position, two of which have been made elsewhere in the literature. The first line of criticism is to say that the Sleeping Beauty problem is one where Beauty's *credence* about the coin toss result is distinct from her *betting behavior* on this matter. The second—which I haven't endorsed—is that the "thirder" position implies a flawed partition of events, as it assumes that the two possible observation times (Monday and Tuesday) can be, for her, distinguishable and mutually exclusive. The last objection, which is my own, argues that the Sleeping Beauty problem should rather be construed as a problem about observation bias, albeit one that involves unnecessary elements about Beauty's location and memories.

Cosmologists often phrase solutions to self-locating problems in terms of "typicality assumptions". In Chapter 4, I distinguished two kinds of typicality assumptions: one assumption concerns the reliability of our data, and the second one concerns their generic character. I have argued that justifications of predictions based on typicality assumptions found in the literature conflate these two kinds of typicality. As a consequence, I claimed that the only typicality assumptions that are warranted (*ceteris paribus*) are those that do not affect our inferences.

This conclusion, in effect, applies more broadly to either of the notions examined in the first three papers: arguments stemming from a state of ignorance or indifference, or from an assumption of typicality, should have no effect on our knowledge of the world if these attitudes are only epistemic in nature, and not the reflection of physical properties.

From these first three papers, one can draw a conclusion that echoes previous critiques of anthropic reasoning. Prior to the anthropic-based prediction from (Weinberg, 1987), we could find the following formulation of the anthropic principle: "we must be prepared to take account of the fact that our location in the universe is *necessarily* privileged to the extent of being compatible with our existence as observers." (Carter, 1974, 293, original emphasis) Weinberg made the same claim when he suggested to use anthropic bounds to restrict the range of possible values of the cosmological constant, but I have shown that what he added to this claim to obtain a prediction—namely, taking into account the number of observers as a function of the value of the cosmological constant—is unwarranted. As a consequence, the only valid element of anthropic reasoning left was already contained in Carter's formulation of the anthropic principle, in which "it is hard to find anything stronger than a tautology." (Earman, 1987)

In the fourth paper, we saw that the rationale behind the appeal to simplicity in model selection is also usually motivated on a priori grounds. Indeed, according to claims made by cosmologists, philosophers, and statisticians, the appeal to simplicity in terms of number of a model's parameters does not depend on our evidence: it is usually argued that, other things being equal, we should prefer models with fewer parameters, because we should avoid overfitting, and the risk of overfitting is greater when we have more, possibly superfluous or meaningless parameters. By examining in greater detail how this claim plays out in cosmology—i.e., with what tools and methods it is applied in model selection—I showed that, strictly speaking, the model selection criteria used or developed do not seek to balance fit-to-data (or predictive accuracy) and "complexity". In accordance with what Norton (2012) showed for the Akaike Information Criterion (AIC), I demonstrated that simplicity in terms of number of free parameters is only a byproduct of the analysis, rather than its aim. The search for simplicity, Nature's modesty, or ontological parsimony is then, at best, no more than an "intuition pump" or a surrogate for background knowledge.

In that same paper, I exposed the specific limitations cosmologists face in many contexts when using statistical and Bayesian model selection criteria. One can find in the literature a critique of using the Bayesian evidence in model selection, as it is highly dependent on choices in the priors, which in many cases in cosmology are poorly motivated (Efstathiou, 2008; van Dyk, 2012). One can also find warnings against the Bayesian Information Criterion if the data sample available to us is small, which can be very small indeed in many contexts in cosmology. I argued, further, that information criteria such as the AIC assume that we have access to multiple instances of the phenomenon or physical system under consideration, which precludes its use when the system in question is unique, such as the universe or the Cosmic Microwave Background radiation.

All the chapters in this dissertation gave examples of probabilistic arguments that lead to spurious confirmation, whether it be because the argument is not valid or because, in these particular instances, the probabilistic framework is not adapted to its object of study. But we also saw that it would be exaggerated to conclude that these are the result of fundamental flaws of Bayesianism. Probabilities can play a legitimate role in cosmological inferences, and the Bayesian framework allows us to avoid unwarranted conclusions.

We saw the risk of substituting a priori reasoning for physical modeling and empirical support. To be sure, a priori assumptions are sometimes useful, and they can even be necessary; for instance, I briefly evoked in § 4.3.4 that we need to assume homogeneity and isotropy at large astrophysical scales in order to be able to work with FLRW models. But I also argued that the Copernican principle's validity (i.e., the "specialness" of our location as observers) can and should be assessed on empirical grounds, and so should the propensity of our universe to foster life, or how many parameters cosmological models should have.

Bibliography

Bostrom, N. (2002a). *Anthropic Bias: Observation Selection Effects in Science and Philosophy*. Routledge.

Bostrom, N. (2002b). Self-Locating Belief in Big Worlds: Cosmology's Missing Link to Observation. *The Journal of Philosophy*, 99(12):607–623.

Bradley, S. (2015). Imprecise Probabilities. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Summer2015 edition.

Carter, B. (1974). Large Number Coincidences and the Anthropic Principle in Cosmology. In *Confrontation of cosmological theories with observational data; Proceedings of the Symposium, Krakow, Poland, September 10-12, 1973*, pages 291–298, Dordrecht. D. Reidel Publishing.

Earman, J. (1987). The SAP Also Rises: A Critical Examination of the Anthropic Principle. *American Philosophical Quarterly*, 24(4):307–317.

Efstathiou, G. (2008). Limitations of Bayesian Evidence applied to cosmology. *Monthly Notices of the Royal Astronomical Society*, 388:1314–1320.

Norton, J. D. (2007). Probability Disassembled. *The British Journal for the Philosophy of Science*, 58(2):141–171.

Norton, J. D. (2008a). Ignorance and Indifference. *Philosophy of Science*, 75(1):45–68.

Norton, J. D. (2008b). The Dome: An Unexpectedly Simple Failure of Determinism. *Philosophy of Science*, 75(5):786–798.

Norton, J. D. (2012). Simplicity in Model Selection. In *The Material Theory of Induction (forthcoming)*, chapter 7.

van Dyk, D. A. (2012). Commentary: Cosmological Bayesian Model Selection. In Feigelson, E. D. and Babu, G. J., editors, *Statistical Challenges in Modern Astronomy V*, Lecture Notes in Statistics, chapter 12, pages 141–146. Springer, New York, NY.

Weinberg, S. (1987). Anthropic Bound on the Cosmological Constant. *Physical Review Letters*, 59(22):2607–2610.

Appendix A

Copyright Permissions

SPRINGER LICENSE
TERMS AND CONDITIONS

Apr 29, 2015

This is a License Agreement between Yann Benétreau-Dupin ("You") and Springer ("Springer") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Springer, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3618380779276
License date	Apr 29, 2015
Licensed content publisher	Springer
Licensed content publication	Synthese
Licensed content title	The Bayesian who knew too much
Licensed content author	Yann Benétreau-Dupin
Licensed content date	Jan 1, 2015
Type of Use	Thesis/Dissertation
Portion	Full text
Number of copies	1
Author of this Springer article	Yes and you are the sole author of the new work
Order reference number	None
Title of your thesis / dissertation	Probabilistic Reasoning in Cosmology
Expected completion date	Oct 2015
Estimated size(pages)	200
Total	0.00 CAD
Terms and Conditions	

Introduction
The publisher for this copyrighted material is Springer Science + Business Media. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

Example 1: Springer Science + Business Media makes no representations or warranties with respect to the licensed material.

Example 2: Springer Science + Business Media makes no representations or warranties with respect to the licensed material and adopts on its own behalf the limitations and disclaimers established by CCC on its behalf in its Billing and Payment terms and conditions for this licensing transaction.

Indemnity
You hereby indemnify and agree to hold harmless Springer Science + Business Media and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

No Transfer of License
This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without Springer Science + Business Media's written permission.

No Amendment Except in Writing
This license may not be amended except in a writing signed by both parties (or, in the case of Springer Science + Business Media, by CCC on Springer Science + Business Media's behalf).

Objection to Contrary Terms
Springer Science + Business Media hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and Springer Science + Business Media (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

Jurisdiction
All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in The Netherlands, in accordance with Dutch law, and to be conducted under the Rules of the 'Netherlands Arbitrage Instituut' (Netherlands Institute of Arbitration).***OR:***

All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in the Federal Republic of Germany, in accordance with German law.

Other terms and conditions:

v1.3

PHILOSOPHY OF SCIENCE ASSOCIATION

PUBLICATION AGREEMENT

Philosophy of Science

THE UNIVERSITY OF CHICAGO PRESS
JOURNALS DIVISION
WWW.JOURNALS.UCHICAGO.EDU

From: [redacted]

To: Yann Benétreau-Dupin

Contribution: Blurring Out Cosmic Puzzles

The Philosophy of Science Association ("the Association"), acting through The University of Chicago Press ("the Press"), is pleased to consider for publication your contribution identified above ("the Contribution"), in its journal, *Philosophy of Science* ("the Journal").

In consideration of the publication of your Contribution, we ask you to assign the copyright to the Association, thus granting us all rights in the Contribution. The Association, in turn, grants to you as author several rights described herein and in the Guidelines for Journal Authors' Rights After Acceptance available on the Web site of the Press (as in effect from time to time, the "Guidelines"). The Press may amend the Guidelines at any time with or without notice to you, and your rights in the Contribution shall be governed by this agreement and the Guidelines in effect at the time of your proposed use. You will receive no monetary compensation from the Association for the assignment of copyright and publication of the Contribution. By signing below, you and the Association agree as follows:

Copyright Assignment:

The Association undertakes to publish the Contribution, subject to approval by the editor(s), in its Journal named above. In consideration of such publication, you hereby grant and assign to the Association the entire copyright in the Contribution, including any and all rights of whatever kind or nature now or hereafter protected by the copyright laws of the United States and of all foreign countries, in all languages and forms of communication. It is understood that the copyright to the Contribution has not been registered with the Library of Congress, but that in the event such registration has taken place you will promptly transfer the copyright registration to the Association. If the Contribution is deemed unacceptable for Publication, you will be notified and all rights will revert to you.

Grant of Rights to Author:

The Association grants to you the following non-exclusive rights, subject to your giving proper credit to the original publication of the Contribution in the Journal, including reproducing the exact copyright notice as it appears in the Journal:

(i) to reprint the Contribution, in whole or in part, in any book, article, or other scholarly work of which you are the author or editor,
(ii) to use the Contribution for teaching purposes in your classes, including making multiple copies for all students, either as individual copies or as part of a printed course pack, provided that these are to be used solely for classes you teach,
(iii) to post a copy of the Contribution on your personal or institutional Web server, provided that the server is non-commercial and there are no charges for access, and
(iv) to deposit a copy of the Contribution in a non-commercial data repository maintained by an institution of which you are a member, after the embargo period identified in the Guidelines and provided all relevant conditions described in the Guidelines have been met.

The rights granted in clauses (i) and (ii) above are intended to benefit the original creators of the Contribution only. Accordingly, if you claim ownership of or rights in the Contribution because it was created by your employee or as a work made for hire, as defined in the Copyright Act, the rights granted in clauses (i) and (ii) above shall not apply to you, and you must contact the Press for permission to make these uses.

Warranties and Indemnifications:

You warrant to the Association as follows:
(i) that the Contribution is your original work;
(ii) that it contains no matter which is defamatory or is otherwise unlawful or which invades rights of privacy or publicity or infringes any proprietary right (including copyright);
(iii) that you have the right to assign the copyright to the Association and that no portion of the copyright to the Contribution has been assigned previously; and
(iv) that the Contribution has not been published elsewhere in whole or in part (except as may be set out in a rider annexed hereto and signed by the Association) and that no agreement to publish is outstanding other than this agreement.

You agree to indemnify and hold the Association harmless against any claim arising from or related to the breach or inaccuracy of any of the warranties listed above.

Requirements for Publication:

You agree to prepare and revise the Contribution according to the instructions of the editor(s) and to meet all other requirements for publication communicated to you by the editor(s) or the Press. It is your responsibility to determine whether the Contribution includes material that requires written permission for publication in the Journal, including any material that is supplementary or ancillary to the Contribution; to obtain such permission, at your own expense, from the copyright owner; and to submit that permission to the editor(s) with the manuscript.

Please complete, sign, and date this agreement, and return it to the editor(s) by mail, fax, or email, retaining a copy for your files. An electronic signature may be used. You agree that an electronic signature shall be valid and binding for all purposes, and hereby waive any objection to use of an electronic version of this agreement as a substitute for the original for any legally recognized purpose. *All joint authors must sign, each on a separate form if necessary.*

AUTHOR SIGN[redacted]

Name: Yann Benétreau-Dupin Date: 04/22/15
(Please print or type)

[redacted]

Telep[redacted] Fax: ____________

Emai[redacted]

FOR THE PHILOSOPHY OF SCIENCE ASSOCIATION:

[redacted]

For internal use only: MSID ____________

Appendix B

Curriculum Vitæ

Name:

Yann Benétreau-Dupin

Post-secondary Education and Degrees:

- B.A./M.A. (Master 1) in Logic, Philosophy, History and Sociology of Science, Université Paris-1 Panthéon-Sorbonne and Università di Bologna, 2007, Hons.
- M.A. in Philosophy, Boston University, September 2010
- Visiting Scholar, University of Pittsburgh, History and Philosophy of Science, Spring 2015
- Ph.D. in Philosophy, Western University (started September 2011)

Selected Scholarships and Awards

- Ontario Trillium Scholarship
  2011–2015
- Rotman Institute Doctoral Entrance Scholarship
  2011
- Graduate Research Assistantships, Rotman Institute of Philosophy
  2012–2014
- Graduate Research Assistantships, C. Smeenk's research funds
  2012–2015
- Rutgers-UCSC Summer Institute in Philosophy of Cosmology Fellowship
  2013
- Richard Hadden award for best student paper presented at the annual meeting of the Canadian Society for History and Philosophy of Science
  2015

Related Work Experience:

- Research and Teaching Assistant (history and philosophy of science for physics teachers)
  Boston University College of Arts and Sciences
  2009–2011

- Administrator
  Boston University Center for Philosophy and History of Science
  2010–2011

- Assistant Editor
  Science & Education
  2010–2014

- Research Assistant (preparation of a course in philosophy of science)
  Boston University School of Education
  Summer 2011

- Teaching Assistant
  The University of Western Ontario
  2011–2013

- Editorial Assistant
  European Journal for Philosophy of Science
  2014–2015

Publications:

Benétreau-Dupin, Y. An Empiricist Criterion of Meaning.
South African Journal of Philosophy 30.2 (2011): 95–108.

Tignanelli, Horacio & **Y. Benétreau-Dupin. Perspectives of History and Philosophy on Teaching Astronomy**.
In Matthews, M.R. (ed.) *International Handbook of Research in History, Philosophy and Science Teaching*, Springer. 2014. 603-640.

Garik, Peter & **Y. Benétreau-Dupin**. **Report on a Boston University Conference December 7–8, 2012 on 'How Can the History and Philosophy of Science Contribute to Contemporary U.S. Science Teaching?'**
Science & Education 23.9 (2014): 1853–1873.

Benétreau-Dupin, Y. & Guillaume Beaulac. **Fair Numbers: What Data Can and Cannot Tell Us About the Underrepresentation of Women in Philosophy**.
Ergo 2.3 (2015): 59–81.

Benétreau-Dupin, Y. Buridan's Solution to the Liar Paradox.
History and Philosophy of Logic 36.1 (2015): 18–28.

Testabilité et signification, translation into French of Carnap, Rudolf (1936–37) "Testability and Meaning." Introduction by Pierre Wagner (IHPST, Paris), Paris: Vrin - Mathesis (2015).

Garik, Peter, L. Garbayo, **Y. Benétreau-Dupin**, C. Winrich, A. Duffy, N. Gross, & M. Jariwala. **Teaching the Conceptual History of Physics to Physics Teachers**, *Science & Education* 24.4 (2015): 387–408.

Benétreau-Dupin, Y. The Bayesian Who Knew Too Much.
Synthese 192.5 (2015): 1527–1542.

Benétreau-Dupin, Y. Blurring Out Cosmic Puzzles.
Philosophy of Science (PSA 2014), forthcoming.